AI in disguise – How AI-generated ads' visual cues shape consumer perception and performance

Yannick Exner[a], Jochen Hartmann[a], Oded Netzer[b], Shunyuan Zhang[c]

[a] *Technical University of Munich, TUM School of Management, 80333 Munich, Germany*
[b] *Columbia Business School, New York, NY 10027, USA*
[c] *Harvard Business School, Boston, MA 02163, USA*

Abstract

Generative AI's recent advancements in creating content have offered vast potential to transform the advertising industry. This research investigates the impact of generative AI-enabled visual ad creation on real-world advertising effectiveness. For this purpose, we collaborate with a display ad platform and leverage a quasi-experimental setting that includes over two million ad-day observations by over seven thousand advertisers across nearly 50 product categories, encompassing more than 16 billion ad impressions and 116 million clicks. We find that display ads in which the image was AI-generated outperform ads with human-generated images in terms of click-through rates, but only if these AI-generated images do not 'look like AI.' We identify key visual features influencing consumers' perception of an ad looking like it was generated by AI. While AI generates more aesthetic images with larger faces, consumers associate these features with human-made ads. In contrast, in line with consumers' expectations, intense color saturation in ads signals AI generation to consumers. Our findings have important implications for advertising platforms that offer AI-powered content creation tools and for advertisers adopting these technologies.
Keywords: generative AI, digital marketing, image analytics, advertising effectiveness

January 14, 2025

1. Introduction

Generative artificial intelligence (AI) plays a transformative role in creative industries, with advertising at the forefront of this transformation (Chui et al. 2023). While it is clear that AI is more cost effective than human labor, the question regarding the ability of AI to generate or assist humans in generating better ads is an open empirical question.

Initial evidence from controlled lab studies (e.g., Miller et al. 2023, Hartmann et al. 2024) and media reports (e.g., Thompson 2024) suggest that AI can generate photorealistic images that are difficult to tell apart from human-made content. Surveys further suggest that humans have mixed reactions to AI-generated content (e.g., Dentsu 2024). But do consumers' stated preferences with respect to generative AI match their revealed preferences in real-world settings?

Answering this question through individual A/B tests is nearly impossible, as these tests only provide a limited view into specific execution or application contexts, which are subject to researchers' discretion and their choice of stimuli. Given the vast variation in the types and quality of ads generated by AI (or humans), any pair of ads chosen for an A/B test can result in superior performance of one type of ads or another.

To address this concern, the present paper takes a bird's-eye view, assessing the effectiveness of AI-generated display ads on consumer behavior across a wide range of product categories, advertisers, and campaign objectives using a large set of actual display ad campaigns generated by real advertisers, where consumer responses to the AI-generated ads directly impact business profits. Hence, our data and analysis have high ecological validity with economic implications (van Heerde et al. 2021). Specifically, we partnered with a leading global online ad platform and obtained large-scale campaign data covering more than two million daily ad-level observations over a period of 73 days. During that period the platform released the GenAI Ad Maker, a free generative AI-enabled tool, allowing advertisers to opt in to co-create ads with the use of generative AI alongside their traditional human-made content.

A limitation of such cross-sectional data is that advertisers may self-select to adopt AI for specific advertising campaigns. To address this limitation, we rigorously define a quasi-experimental design within our observed data. We utilize the fact that advertisers routinely employ an experimental mindset by creating ad variations within the same campaign to 'A/B test' the effectiveness of different ad designs. Our quasi-experimental setting leverages these naturally occurring experiments. Specifically, we focus on ads that were created by the same advertiser on the same day and as part of the same campaign, sharing the same landing page, promoting the same product with the same campaign objective, but included design variations with both AI-generated and human-made images. This design allows us to use experiment-specific fixed effects to isolate the effect of AI-generated images on ad effectiveness and mitigate potential confounders arising from advertisers' decisions to adopt the GenAI Ad Maker in general or use it only for specific campaigns, thus ensuring consistent advertiser characteristics, identical temporal context, and uniform campaign strategies. Although advertisers can use the GenAI Ad Maker to generate both images and captions, this paper specifically focuses on AI-generated imagery. This is because image content, as compared to text (caption), is more difficult for advertisers to modify and thus allows for a cleaner identification of the treatment.

Overall, our analysis of the ads' click-through rates (CTR) in our quasi-experimental setting reveals that, on average, AI-generated ads perform comparable to human-made ads, while being orders of magnitude cheaper to create (Hartmann et al. 2024, Reisenbichler et al. 2022).

However, while these results suggest parity between AI-generated and human-made ads at the aggregate level, they may mask important heterogeneity in consumer response to AI outputs. Drawing on "algorithm aversion" theory (Dietvorst et al. 2015, Castelo et al. 2019), which posits that humans tend to prefer humans over AI algorithms, we propose that consumer reactions to ads may vary based on the perceived artificiality of the AI-generated content. Importantly, in our setting, AI generation is not disclosed to consumers. But what if consumers can suspect that an ad is AI-generated even without disclosure? If certain image characteristics make ads appear AI-generated, will consumers, consciously or not, infer the ads' origin and react negatively to such ads? If consumers' aversion to AI plays a role in consumers' response, then AI-generated ads might underperform human-made ads if they appear AI-generated and possibly outperform human-made ads if they appear human-made.

To explore this question empirically, we introduce a looks-like-AI measure that captures humans' perception on whether they perceive an ad's creative as AI-generated or human-made, i.e., an ad's perceived artificiality (Jakesch et al. 2023). Interestingly, we find that more than 45% of the AI ads from our quasi-experimental data are perceived as definitely or likely human-made, indicating that AI-generated images can disguise their origin from human observers. This finding is consistent with anecdotal evidence, suggesting that people cannot reliably detect AI-generated images (Thompson 2024). We introduce perceived artificiality into our quasi-experimental analysis and find that AI-generated images that do not look like AI significantly outperform human-made ads' CTR. In contrast, AI-generated images do not yield higher CTRs if humans perceive them as AI-like.

This raises the question of how AI-images differ from human-made ones, and more importantly, what makes consumers perceive an ad as AI-generated. Exploring a broad range of perceptual, structural, and content-related image dimensions (Hartmann et al. 2024), we find that, compared to human-made images, AI-generated images tend to be more colorful, feature larger faces, and have higher aesthetics scores. Although consumers perceive ads with vivid colors to be AI-generated, larger faces and higher image aesthetics are negatively related to consumer perception that an ad is AI-generated. Thus, larger displayed faces, which are more common in AI-generated ads than human-made ads can help disguise AI ads, and increase their trustworthiness (Nightingale and Farid 2022), corresponding to a higher CTR.

Our paper makes three important contributions. First, we demonstrate the effectiveness of AI-generated online ads in a real-world context. In a quasi-experimental setting, we show that ads with AI-generated images achieve human-level CTRs. Second, we identify the boundary condition of an ad's perceived artificiality. If AI-generated images do not look like AI, they can achieve superhuman CTRs, offering evidence-based recommendations for optimizing the use of AI in visual marketing. Thus, we add to the literature on human perception of machine-made outputs, specifically AI-generated ads, by identifying key visual features that consumers associate with AI-generated images. Third, we offer insights into advertisers' adoption of generative AI tools to enhance their online ad performance (see Reisenbichler et al. 2022), providing important practical implications for both platforms and advertisers to enhance their ad effectiveness and efficiency.

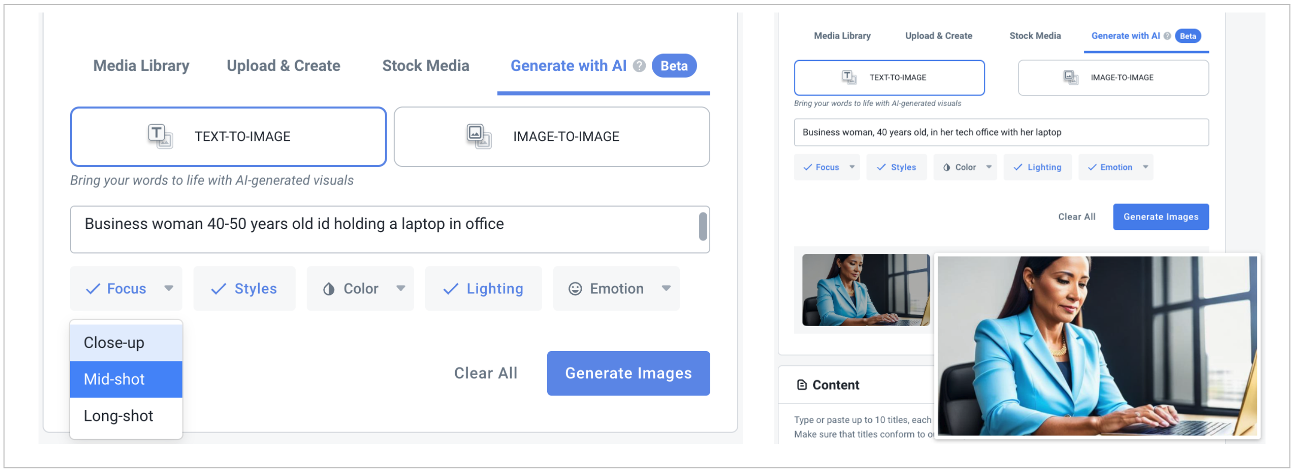
2. Background and data

We collaborated with Taboola, a major US-based digital advertising platform, which publishes advertiser-provided ads across publisher websites like MSN, NBC News, or USA TODAY, reaching 500 million daily active users (Feeney 2023).

In July 2023, the platform released the GenAI Ad Maker to all English advertisers (Feeney 2023) following a limited beta test that began in February, 2023 (Taboola 2023). The launch of the GenAI Ad Maker marked a shift for the platform from its traditional ad publishing role of ad allocation to a more extensive role, allowing advertisers to "create more efficient and effective ads at scale while saving valuable time and resources" (Taboola 2023).

The GenAI Ad Maker is integrated into an advertiser's workflow when setting up campaigns. Advertisers can either upload their own visuals and captions or generate them within the platform free of charge. If advertisers choose to use the GenAI Ad Maker, the AI creates multiple ad variations for advertisers to choose from, based on their written input (i.e., the prompts). Stable Diffusion 2 is used to generate images while OpenAI's GPT 3.5 generates ad captions (Feeney 2023). Figure 1 presents the advertiser interface of the GenAI Ad Maker.

Figure 1: The GenAI Ad Maker user interface as seen by advertisers



Typical of display ads, each ad in our data is a unique combination of an image (ad creative) and a textual caption (ad copy). Our study focuses specifically on ads that use the GenAI Ad Maker for image creation rather than solely for text creation. This focus is motivated by three key factors. First, while some evidence exists for the effectiveness of AI-generated text in advertising (e.g., Reisenbichler et al. 2022), there is a notable gap in understanding the real-world effectiveness of AI-generated marketing imagery. Second, modifying AI-generated text is relatively straightforward (Jakesch et al. 2023), often involving simple word or letter replacements. In contrast, modifying AI-generated images typically requires advanced editing skills and tools (e.g., Photoshop), presenting a higher barrier. This distinction allows for a more reliable identification of whether the advertiser directly used the AI-generated images, enhancing the internal validity of our analysis (see Section 2 for detailed discussions). Third, compared to texts, images have a well-documented positive and significant influence on consumer engagement (Li and Xie 2020), a phenomenon known as "picture superiority" (e.g., Paivio and Csapo 1973), making the study of AI-generated images especially valuable for understanding overall ad effectiveness. That being said, we control in our analyses for AI-generated text as well.

Our data comprises all active English ads on the ad platform from 6/3/2023 to 8/15/2023. The platform released the GenAI Ad Maker during our data collection window – on 7/12/2023.

The data contains daily metrics for each ad, including impressions, clicks, and associated spend. Each ad contains the caption, image, description, and its associated campaign ID and advertiser ID. For each campaign, the data includes its marketing objective (i.e., campaign objective) as well as an advertiser type and product category. Our dataset covers 305,121 ads with an average runtime of 7.30 days, resulting in 2,227,664 ad-day observations by a total of 7,074 advertisers. On average, each advertiser ran 43 ads across 6 campaigns during our observation period. Overall, our data include more than 16.4 billion impressions and over 116 million clicks with an average CTR of .71%.

By the end of the observation period, 6.25% of the advertisers used the GenAI Ad Maker at least once. Across all ads, the advertisers used the AI tool to generate 3.04% and 2.43% of the images and captions, respectively, after the GenAI Ad Maker went live [1].

2.1. A quasi-experimental setting

Despite the appeal of our large-scale, real-world dataset, which offers rich insights into the actual adoption of AI by advertisers and real consumer reactions to ads, working with real-world observational data comes with inherent challenges:

(i) Advertisers self-select into using the GenAI Ad Maker.

(ii) Even within an advertiser, advertisers may decide which campaigns to leverage the GenAI Ad Maker based on unobserved expectations about their ads' performance.

(iii) Advertisers may start ads on different dates, which can lead to variation in consumer response due to unobserved external events (e.g., news).

(iv) Advertisers may keep more effective ads in the field for a longer period of time, which may result in more ad-day observations for effective ads versus less effective ads.

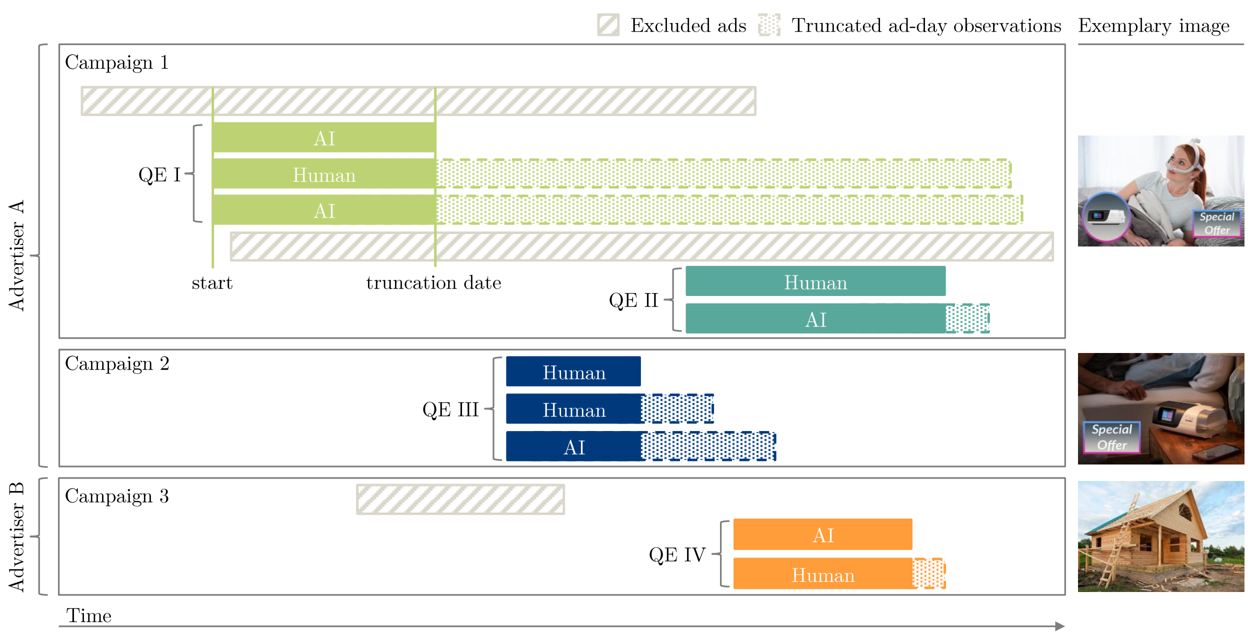
[1] Our data does not include clickstream of the advertiser interaction with the tool. We only observe full adoption of the GenAI Ad Maker's generated output. This means we only observe if advertisers used AI-generated content as proposed by the GenAI Ad Maker without further modifications.

(v) Advertisers may manually upload AI-generated content to the platform which we cannot observe.

(vi) The platform's targeting algorithm might also influence the ad effectiveness.

To address these challenges, we construct a quasi-experimental setting (see Figure 2 for a visual schema). Typically, advertisers run multiple ads in a single campaign to test different ad variations with an experimental mindset. These 'sibling ads' share the same advertiser, landing page, and cater towards the same campaign objective and product. We leverage such naturally occurring 'experiments' but need to account for the challenges mentioned above as follows.

Figure 2: Schema of the quasi-experimental setting



Notes. Dashed bars indicate ad-day observations excluded because of right truncation; gray bars indicate ads excluded as only a single ad, or multiple ads that use purely AI-generated or human-made images; QE indicates a quasi-experiment.

To address challenges (i) and (ii), we only assess campaigns that include both ads with AI-generated and human-made images.[2] This design allows us to include a quasi-experiment

[2]We confirm that these human-made ads do not differ significantly from those not included in our quasi-experiments in terms of our dependent variable, CTR (see Web Appendix Appendix E.1).

fixed effect to explore within campaign and within advertiser variation as there might be multiple quasi-experiments within the same campaign that start on different dates. Thus, we mitigate potential concerns regarding endogenous decisions by the advertisers to adopt the GenAI Ad Maker in general or for specific campaigns. Further, we require all ads in a quasi-experiment to be created on the same day, addressing challenge (iii).

For example, campaign 1 in Figure 2 features two distinct quasi-experiments as the ads in each quasi-experiment feature different starting days and two of the dates included only a single ad design (first and fifth bar). To ensure comparable runtimes and address a potential 'survivor bias' as defined in challenge (iv), we truncate all ads once the first ad in a quasi-experiment stops (see 'truncation date' dotted lines in Figure 2 which indicate the ad-day observations we truncate). As a result, the constructed approach mimics an A/B test: within each quasi-experiment the ads in the treatment (AI-generated) and control (human-made) groups are identical across various dimensions (time, campaign, landing page, campaign objective, product, etc.). This quasi-experimental approach addresses challenges (i) to (iv).

While our data limit our ability to fully address challenges (v) and (vi) directly, we have taken steps to mitigate these concerns and gathered indirect evidence suggesting that they are unlikely to significantly undermine our main findings. For challenge (v), we cannot detect if advertisers manually upload AI-generated content. If some human-labeled content is actually AI-generated, and if AI generally outperforms human-made content, then our analysis underestimates AI's true performance advantage. In other words, any undetected AI content in our human-made control group would make it harder to find significant performance differences between AI and humans, resulting in a conservative estimate of the AI ads' performance. Moreover, interviews with the platform's experts and advertisers confirm this practice to be rare due to its inefficiency versus using the platform's seamlessly integrated GenAI Ad Maker.

Regarding challenge (vi), observing ad performance within the platform's targeted environment reflects an inherent reality of online marketing — the real world operates through targeting algorithms that determine ad placement, frequency, and audience (Boegershausen et al. 2025, Braun et al. 2024). As such, researchers can only evaluate ad effectiveness as “the combined impact of advertising creative elements and algorithmic targeting” (Braun et al. 2024). Thus, to the extent that the algorithm differentially targets ads within our quasi-experimental design, the result could be interpreted as the combined differential impact of AI-generated ads relative to human-made ads, due to creative elements and algorithmic targeting. To address the question of whether generative AI ads perform better or worse than human-made ads, the combined effect is arguably the right measure. That being said, due to its relatively limited first-party data, the Taboola platform algorithmic targeting is less of a concern relative to platforms like Meta. Additionally, we find that the total exposure of ads in terms of logarithmic mean daily impressions did not differ between the experimental groups, mitigating our concern about algorithmic differences driving the observed effects ($\beta_{AI-generated\ image} = -.3152$, $SE = .1676$, $p = .0602$; $\beta_{AI-generated\ caption} = -.0609$, $SE = .2869$, $p = .8320$).

Focusing on our quasi-experimental setting, our eventual dataset includes 4,633 ads across 1,186 quasi-experiments within 351 campaigns run by 202 advertisers. On average, each quasi-experiment includes 3.91 ads, 25.66% of which are AI-generated images. The most common product categories in our datasets are 'technology and computing', 'personal finance', and 'home and garden' (22.75%, 16.02%, and 13.84%, respectively) while the most common campaign objectives are 'leads', 'purchases', and 'page views' (52.32%, 33.67%, and 7.17%, respectively). See Web Appendix Appendix A for further details.

3. Consumer reaction to AI-generated ads

Using the constructed quasi-experiments from our real-world data, we compare AI-generated images with human-made images to answer our research question: Can AI-generated content match or even outperform human-made ads?

Our main outcome variable is CTR, measured as clicks divided by impressions, a commonly used measure to assess ad performance (Robertson et al. 2023, Boegershausen et al. 2025). Model-free evidence suggests that AI-generated ads generate an average CTR of .76% (885,434 clicks, 117,002,984 impressions) while human-made ads generate a CTR of .65% (1,639,340 clicks, 252,530,342 impressions), indicating that AI-generated ads may outperform human-made visual marketing content ($\chi^2(1, N = 369{,}533{,}326) = 13{,}641$, $p < .001$). However, these results do not account for possible within-campaign differences.

Our empirical model to assess the impact of ad generation type on consumer response to an ad follows Robertson et al. (2023):

$$clicks_{ijt} \sim Binomial(impressions_{ijt}, \theta_{ijt}) \tag{1}$$

where $i = 1, ..., N$ indicates ad i in quasi-experiment $j = 1, ..., J$, and $t = 1, ...T$ reflects the calendar date an observation was recorded. Accordingly, θ_{ijt} refers to the CTR of ad i in quasi-experiment j for a given day t. The number of clicks on the ad follows a binomial distribution where $\theta \in [0, 1]$ is the probability of a consumer impressed with ad i from quasi-experiment j on day t clicking on the ad in a single Bernoulli trial. We estimate the effect of AI-generated content on an ad's CTR using the following regression model[3]:

$$\text{logit}(\theta_{ijt}) = \beta_0 + \beta_1 x_{ij}^{image} + \beta_2 x_{ij}^{caption} + \gamma X_{ijt} + \alpha_j + \delta_t \tag{2}$$

where x_{ij}^{image} and $x_{ij}^{caption}$ are binary variables indicating if ad i in quasi-experiment j uses an AI-generated image or caption, respectively, where zero indicates a human-made ad. Our primary focus is on the coefficient β_1. This coefficient represents the change in log odds of ad i's CTR in quasi-experiment j on day t when using an AI-generated image relative to a human uploaded image. We control through $x_{ij}^{caption}$ for whether a caption was generated by AI or a human. X_{ijt} is a matrix of control variables including a set of verbal features

[3]In the binomial model, observations are weighted by their respective number of impressions to account for differences in sample size (i.e., impressions) across ad-day observations.

Table 1: Performance of AI-generated images

Dependent Variable:	CTR	
Model:	(1)	(2)
Variables		
AI-generated image	-.0730	-.0397
	(.0692)	(.0490)
AI-generated caption	-.0631	-.1199
	(.1761)	(.2099)
Controls		
CPM		Yes
Verbal features of caption		Yes
Verbal features of description		Yes
Visual features of creative		Yes
Fixed effects		
Quasi-experiment ID	Yes	Yes
Calendar date	Yes	Yes
Fit statistics		
Observations	29,592	29,592
Squared Correlation	.1938	.4118
Pseudo R^2	.9823	.9847
BIC	547,157	474,485

$^{***}: p < .001, ^{**}: p < .01, ^{*}: p < .05, \dagger: p < .1$

Notes. Clustered standard errors at the quasi-experiment level in parentheses. Continuous independent controls are standardized and mean-centered. Observation count differs slightly from 29,631 reported in Web Appendix Appendix A due to lack of variation in the DV for 39 (0.13%) observations.

in the ads caption and description (e.g., word count or authenticity), visual features of an ad's image (e.g., contrast or aesthetics), and the ad's cost per mille (CPM). Web Appendix Appendix B and Appendix C list all verbal and visual control variables used for the creatives and captions/descriptions, respectively. α_j and δ_t represent fixed effects for quasi-experiment j and calendar date t, respectively. All continuous independent variables are z-standardized for increased interpretability.

Model 1 of Table 1 presents the results from estimating the model without controls γX_{ijt}. As can be seen from Model 1, AI-generated images perform insignificantly different from human-made images in terms of their CTR ($\beta_{AI-generated\,image} = -.0730$, $SE = .0692$, $p = .2914$). In Model 2, the results remain consistent when we add the controls γX_{ijt} ($\beta_{AI-generated\,image} = -.0397$, $SE = .0490$, $p = .4173$).

First, note that these results suggest that AI-generated images in ads, which are likely more cost-efficient, elicit a similar response from consumers as human-made images in ads. This highlights the potential for generative AI technologies to democratize the ad creation while maintaining reasonable performance.

Furthermore, while these results suggest parity between AI-generated and human-made ads at the aggregate level, they may mask important heterogeneity in consumer response to AI-generated content. In particular, since consumers might exhibit negative predisposition against AI-generated content (Castelo et al. 2019, Horton et al. 2023), it is plausible to expect that consumers react less favorably to AI if they can tell that an ad was AI-generated. Can consumers infer the generation of an image as AI-generated or human-made and then potentially discriminate upon it? If consumers' aversion to AI plays a role in driving the lack of an average treatment effect, then will AI-generated ads outperform human-made ads if they do not look like AI? To address this question, we collect additional data on people's perceptions of the ads.

4. AI-generated ads' perceived artificiality

We introduce a variable we term 'looks-like-AI' that captures humans' perception of an image as AI-generated (i.e., perceived artificiality[4]).

For all unique images in our quasi-experiments (N = 1,751 with N = 460 AI-generated and N = 1,291 human-made images), we ask 5 MTurkers per image to rate it on a 5-point Likert scale: 'Is this image human-made OR AI-generated?' (adopted from Jakesch et al. 2023). We normalize the ratings on perceived artificiality $\in [0, 1]$ (see Web Appendix Appendix D for further details).

Figure 3 displays the distribution of perceived artificiality. While AI-generated images ($M = .4891$, $SD = .2842$) achieve a significantly higher artificiality score than human-made ones ($M = .3848$, $SD = .2581$; $t(745) = 6.9235$, $p < .001$), humans struggle to clearly

[4] Hereafter, we use interchangeably 'perceived artificiality' and 'looks-like-AI'.

identify AI-generated from human-made images in real-world advertising. Raters identified 24.87% of the human-made ad images as likely or definitely AI-generated, and 58.92% of the AI-generated ads, as not sure, likely, or definitely human-generated.

Figure 3: Perceptual human ratings on ads' perceived artificiality

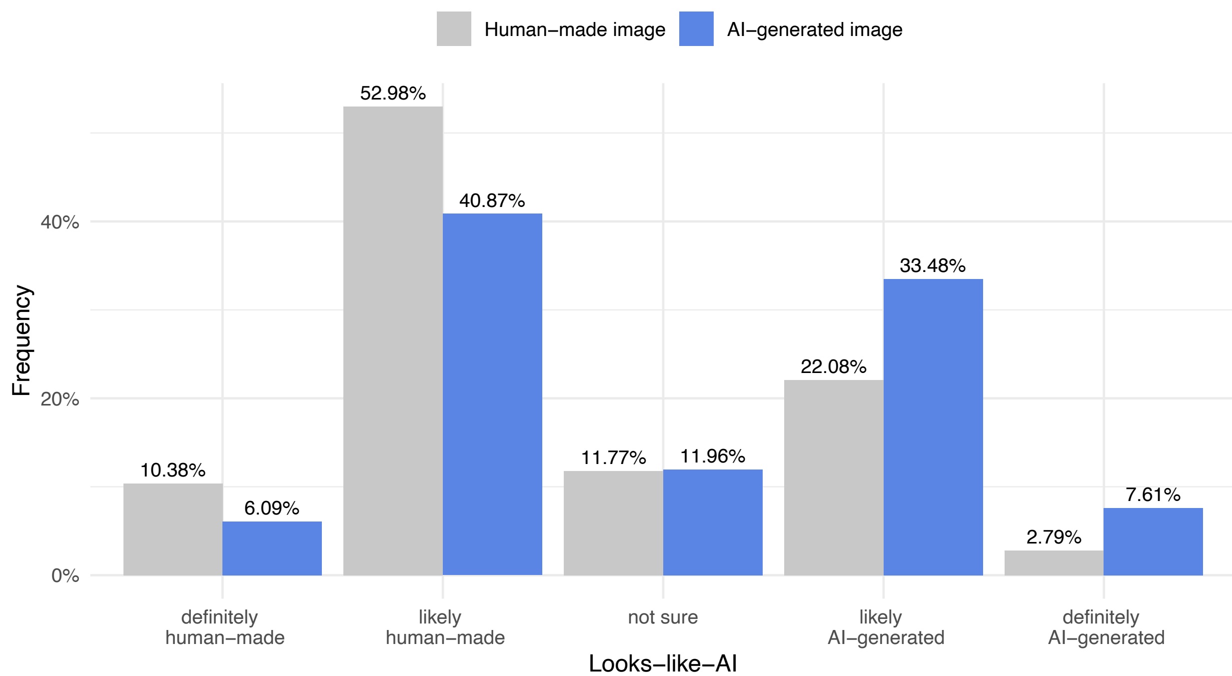
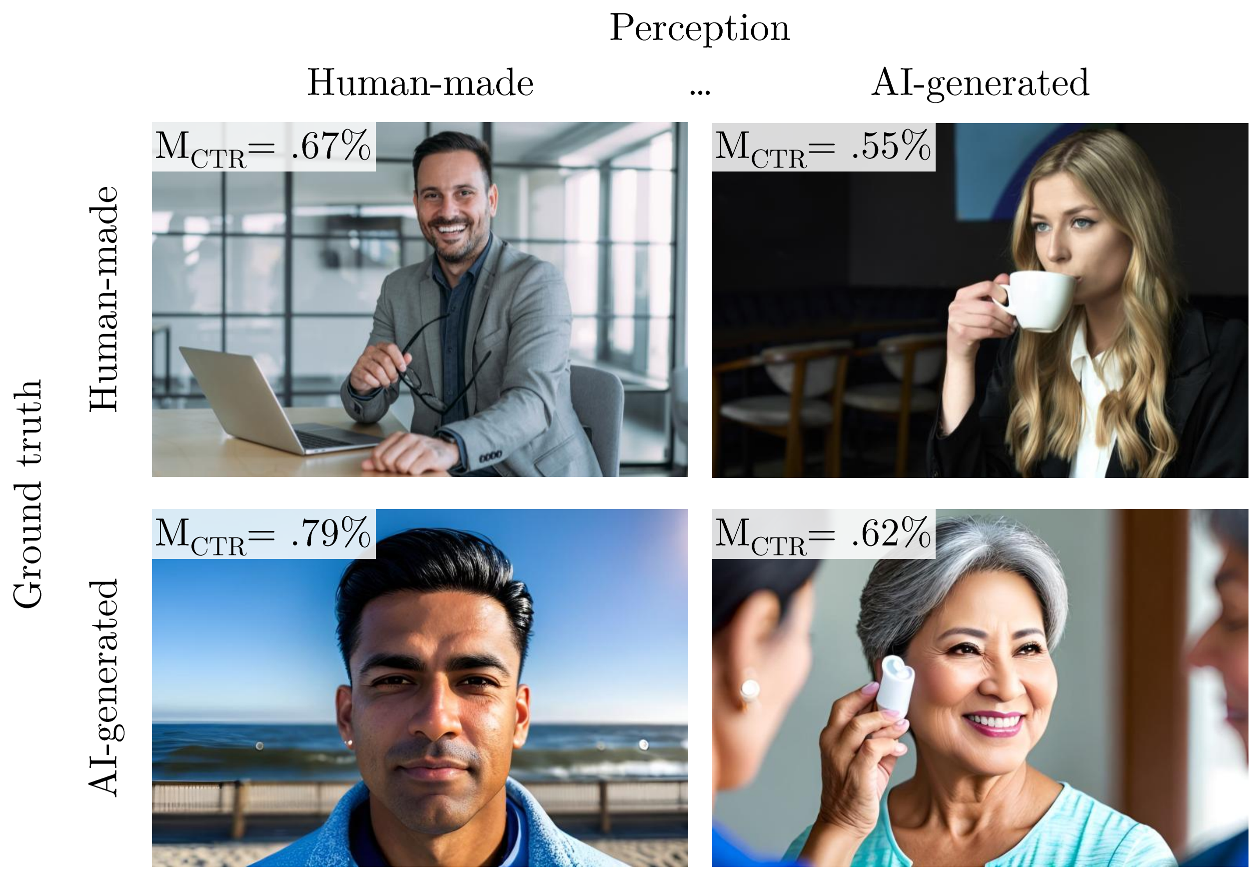


Figure 4 illustrates a selection of images from our data, juxtaposing AI-generated and human-made content as well as their perceived artificiality alongside average CTRs observed across all ads that belong to a cell. Model-free, AI-generated images that disguise their origin (i.e., are rated as 'definitely human-made,' 'likely human-made,' or 'not sure') achieve a mean CTR of .79% outperforming the mean .62% CTR of AI-generated images that do not disguise their origin (i.e., are perceived as 'likely' or 'definitely' AI-generated), as well as human-generated ads that look or do not look like AI (.55% and .67%, respectively).

Figure 4: 2 × 2 of ad generation sources and ad perception with exemplary images, and average CTR per cell



To explore potential differences in consumer reactions to AI-generated versus human-made content, we extend our regression analysis by incorporating the looks-like-AI measure. This allows us to test whether consumers' CTRs vary based on their perception of who created an ad (i.e., human or machine) and based on the degree of an ad's perceived artificiality. Specifically, we replicate the regressions from Table 1, with looks-like-AI as both main effect (Models 1 and 2) and interaction with AI-generated image (Model 3). We present the estimation results in Table 2.

Consistent with the literature on AI aversion, we find that consumers' perception of content as AI-generated has a strong negative effect on CTR ($\beta_{looks-like-AI} = -.3759, SE = .0833, p < .001$; see Table 2 Model 1). This effect persists even after controlling for visual features of the image, verbal features of both caption and description, and CPM

($\beta_{looks-like-AI} = -.3468, SE = .1159, p < .01$; see Table 2 Model 2). To further disentangle this effect and understand whether consumers react differently to actual AI-generated content that appears artificial, we examine the interaction between AI generation and perceived artificiality.

Notably, for 'definitely human-made'-looking images (recall looks-like-AI is standardized in our regression from $\in [0, 1]$, where 0 represents 'definitely human-made'), AI-generated images have a positive effect on an ad's CTR compared to human-made ads ($\beta_{AI-generated\ image} = .2872$, $SE = .0995$, $p < .01$), but the more an AI-generated image looks like AI, the lower its CTR ($\beta_{AI-generated\ image \times looks-like-AI} = -.5175$, $SE = .1764$, $p < .01$; see Table 2 Model 3 and Web Appendix Figure D.4 for a bar plot with predicted CTR values). This indicates that consumers particularly penalize AI-generated content when they sense it might be AI-generated, exhibiting a stronger negative reaction to perceived artificiality in AI-generated versus human-made content (Dietvorst et al. 2015).

4.1. Robustness Analyses

We conduct a broad set of robustness analyses to assess the empirical validity of our results and find consistent results as detailed in Web Appendix Appendix E. Specifically:

(1) To assess for possible self-selection in quasi-experiments that involve AI-generated images, we compare the CTR of human-made ads included in our quasi-experimental sample with those in campaigns that included only human-generated images and find no significant differences.

(2) Our results are robust to excluding short-running ads, where advertisers may have simply 'played around' with the GenAI Ad Maker.

(3) Our results are robust when excluding ads with very few impressions or very high CTRs, respectively.

Table 2: CTR performance of AI-generated images depending on consumers' perceived artificiality (looks-like-AI)

Dependent Variable:		CTR	
Model:	(1)	(2)	(3)
Variables			
AI-generated image	-.0242	.0430	.2872**
	(.0562)	(.0549)	(.0995)
AI-generated caption	-.0762	-.1127	-.1090
	(.1803)	(.2152)	(.2153)
looks-like-AI	-.3759***	-.3468**	-.1735
	(.0833)	(.1159)	(.1365)
AI-generated image × looks-like-AI			-.5175**
			(.1764)
Controls			
CPM		Yes	Yes
Verbal features of caption		Yes	Yes
Verbal features of description		Yes	Yes
Visual features of creative		Yes	Yes
Fixed effects			
Quasi-experiment ID	Yes	Yes	Yes
Calendar date	Yes	Yes	Yes
Fit statistics			
Observations	29,592	29,592	29,592
Squared correlation	.1916	.4126	.4134
Pseudo R^2	.9825	.9848	.9849
BIC	540,161	471,317	469,581

$^{***}: p < .001, ^{**}: p < .01, ^{*}: p < .05, \dagger : p < .1$

Notes. Clustered standard errors at the quasi-experiment level in parentheses. Continuous independent controls are standardized and mean-centered.

(4) Our results are robust when we focus only on the first few days of a campaign, a time frame in which the allocation algorithm has gathered only limited information about an ad for targeting purposes (Boegershausen et al. 2025).

(5) We get consistent results when excluding pre-go-live ads and advertisers who participated in the GenAI Ad Maker's beta phase.

4.2. Which visual features shape consumer perception of an image to look like AI?

Having established the impact of AI-generated images on ad performance, we turn our attention to the features that drive consumers' perception and how AI text-to-image models fare on these features. Understanding these deep-seated perceptual mechanisms is crucial, as Tables 1 and 2 highlight that AI ads are effective only if an ad's image does not look like AI. While generative AI models will continue to evolve, the fundamental question remains: What visual features drive consumers' perception of artificiality?

Prior literature suggests a rich set of visual features that shape how consumers react to visual stimuli (e.g., Zhang et al. 2022, Talebi and Milanfar 2018, Li and Xie 2020). Building on established theoretical frameworks of visual perception, we examine visual features that relate to perceptual (aesthetics, quality, realism), structural (e.g., color saturation), and content-related (e.g., includes text) aspects that influence how consumers process and evaluate images (Hartmann et al. 2024). Web Appendix Appendix C describes these 18 fundamental features in more detail.

We regress the looks-like-AI score of each ad image present in our quasi-experiments, on the visual features to explore the determinants of looks-like-AI to assess which visual features drive consumers' perceived artificiality. Table 3 Model 1 presents the corresponding results. We z-transform all continuous feature values for ease of interpretation. We find that consumers perceive images with higher feature values for color, kurtosis, and warmth as rather AI-generated. Likewise, images which display text are perceived as more likely to be AI-generated. This finding is plausible given that generative text-to-image diffusion models often render text incorrectly, a signal that consumers are likely using to assess perceived artificiality, see Web Appendix Figure C.2 for an example.

In contrast, people perceive images with high aesthetics and facial presence (specifically medium to large faces) as human-made, being associated with lower perceived artificiality. This finding aligns with Miller et al. (2023) who find that humans can perceive AI-generated faces as "more real than human ones."

Next, we explore which visual features are actually more likely to occur in AI-generated ad images versus human-made ones and whether people correctly attribute features that often appear in AI-generated images to perceived artificiality.

4.3. Which visual features are actually present in AI-generated ads?

To contrast the visual features that drive human perceptions with the features that actually appear in AI-generated images, we assess which features are more or less pronounced in AI-generated versus human-made imagery. We use a binary logit regression with a dummy variable indicating whether an image is AI-generated as the dependent variable and the same visual features used in the previous subsection as independent variables. See Table 3, Model 2.[5]

The results show meaningful differences between the GenAI Ad Maker's AI-generated and human-made images with respect to visual features. Specifically, AI-generated images are significantly more aesthetic, more blurry, more clear, more colorful, feature more contrast, are warmer, and more frequently sports-themed. Further, AI-generated images are significantly less likely to display text. This finding is in line with common difficulties that AI-generated images face when it comes to displaying text (Hartmann et al. 2024). We further find that AI-generated images are less bright, sharp, and skewed compared to human-made images. Interestingly, AI-generated images create larger faces. Compared to human-designed ads, they are significantly less likely to generate small and medium-sized faces but directionally

[5] Web Appendix Figure C.1 presents model-free evidence of the prevalence of different visual features in AI and human-made images.

Table 3: Visual features associated with perceived artificiality, and with images actually generated by AI

Dependent Variables:	looks-like-AI	is-AI-generated
Model:	(1)	(2)
	OLS	Logistic Regression
Variables		
Perceptual		
Aesthetics	-.0911**	.5592***
	(.0295)	(.1154)
Quality	.0139	.0078
	(.0245)	(.0760)
Realism	-.0777**	-.0605
	(.0247)	(.0787)
Structural		
Blur	-.0312	.6194***
	(.0314)	(.1119)
Brightness	.0517	-1.651***
	(.0463)	(.2491)
Clearness	-.0407	.9201***
	(.0381)	(.1538)
Color	.1313***	.7171***
	(.0291)	(.1079)
Contrast	.0407	1.341***
	(.0294)	(.1314)
Kurtosis	.0866**	-.3555
	(.0287)	(.3637)
Sharpness	.0590†	-2.278***
	(.0317)	(.4603)
Skewness	.0101	-1.581***
	(.0436)	(.3194)
Symmetry	.0331	-.0859
	(.0258)	(.1003)
Warmth	.0687**	.2727***
	(.0242)	(.0798)
Content		
Face size small	-.1797**	-.5161*
	(.0659)	(.2283)
Face size medium	-.2709***	-.4610*
	(.0711)	(.2323)
Face size large	-.2218**	.4447†
	(.0765)	(.2358)
Food-themed	-.2406†	.5319
	(.1240)	(.4199)
Sports-themed	.0500	.7630**
	(.0897)	(.2683)
Technology-themed	-.0639	.1233
	(.0664)	(.2005)
Text displayed	.1478*	-1.194***
	(.0724)	(.3018)
Controls		
AI-generated image	.4453***	
	(.0657)	
Fit statistics		
Observations	1,751	1,751
Squared correlation	.1230	.4755
R^{2a}	.1123	.4264
BIC	4,903	1,314

$^{***}: p < .001, ^{**}: p < .01, ^{*}: p < .05, \dagger: p < .1$

Notes. Robust standard errors in parentheses. Continuous independent controls are standardized and mean-centered. Constant omitted from table. We clustered facial area into three equi-sized categorical bins (reference: no face). [a] refers to adjusted R^2 for OLS regression in Model 1 and pseudo R^2 for logistic regression in Model 2.

more likely to generate large faces.

4.4. AI in disguise – Comparing human perception with AI reality

While prior research shows that consumers often cannot accurately identify AI-generated content (Miller et al. 2023, Hartmann et al. 2024), our analysis reveals both alignment and misalignment between perception and reality for visual features in AI-generated images. Interestingly, we find a small negative, though insignificant, correlation between the features most present in AI-generated images versus those that make an image look like AI (the correlation between the coefficients in Models 1 and 2 from Table 3 is $\rho = -.1851$, $p = .4348$).

In some cases, consumer intuition proves in line with AI's behavior. For example, the more colorful and warm an image is, the more likely consumers are to identify it as AI-generated, matching actual characteristics of AI-generated images.

However, AI effectively disguises itself through other features. Specifically, AI generates images with higher aesthetics scores and larger faces – features that consumers typically associate with human-made content. Similarly, in rare cases where AI successfully renders text, it tends to deceive consumers who associate text with human creation. These findings demonstrate how AI can leverage misaligned perceptions, particularly through larger face size and higher aesthetics, to appear more human-made than it actually is.

5. Which advertisers benefit most from AI?

Next, we assess whether certain advertisers benefit more than others from the GenAI Ad Maker. Recall that our quasi-experimental design controls for time-invariant advertiser heterogeneity via the quasi-experiment fixed effects, which are nested within advertisers. Hence, our main analysis capitalizes on within-advertiser variation. However, there may be differences across advertisers in terms of how strongly they benefit from AI-generated images in their campaigns. To examine such heterogeneity, we interact the effect of an AI-generated image with two important advertiser characteristics (see Web Appendix Table E.11 for details). First, we assess whether the benefit of using AI-generated images in advertising varies based on scale and experience of the advertiser proxied by their cumulative impressions of the advertiser's ads. We find that smaller and less experienced advertisers benefit relatively more, indicating an asymmetric distribution of the value created by AI-generated images (See Web Appendix Table E.11, Model 1 for details). This result is consistent with findings in the literature (e.g., Noy and Zhang 2023), that generative AI technologies level the playing field between players of different experiences.

Second, we assess the impact of AI-generated images for ad arbitrage advertisers. Arbitrage advertisers operate on a volume-based business model where they use ads to drive traffic to content pages and then monetize that traffic by displaying (more profitable) ads to these visitors. Our findings (see Web Appendix Table E.11, Model 2) show that ad arbitrage advertisers are less likely to benefit from AI-generated images in their campaigns versus regular advertisers. In fact, they do better if they stick to human-made images, as they achieve significantly lower CTRs when using AI to design an ad image. Presumably, this pattern could be explained by the fact that these advertisers have already optimized their human-made ad creation process with established automated workflows for creating and testing ads. Given their expertise in arbitrage-specific advertising, they may not have yet developed equivalent proficiency in optimizing AI tools to match their unique business requirements shortly after the release of the GenAI Ad Maker.

6. General Discussion

Drawing on carefully constructed quasi-experiments from real-world data encompassing over 16 billion ad impressions, we find nuanced patterns in consumers' response to AI-generated advertising visuals: AI performs best if it disguises itself. On average, AI-generated ads perform insignificantly differently from human-made ads. However, this aggregate effect masks important heterogeneity in terms of consumer response to AI-generated outputs. Specifically, we find that AI-generated visual marketing content outperforms human-made images if the content does not look like AI.

Our research has implications for both theory and practice. First, we demonstrate real-world empirical evidence on the effectiveness of AI-generated online ads. Given the tremendous efficiency gains that come along with AI-generated content (e.g., Reisenbichler et al. 2022, Noy and Zhang 2023, Hartmann et al. 2024), adoption of these tools is likely to continue to increase rapidly. However, as our findings suggest, efficiency gains should not come at the cost of effectiveness losses. We contribute to the understanding of how AI-generated content can outperform human-made images in effectiveness as well. Importantly, we do not only show that AI works, but also when it does *not* work, identifying ads' perceived artificiality as an important boundary condition: AI-generated images achieve superhuman CTR levels only when they do not appear to be AI-generated.

Second, we add to the literature on human perception of AI-generated visual ads. We identify key visual features that drive consumers' perception of images as AI-generated or human-made and contrast these to the reality of AI-generated content. For some visual features like color and warmth, consumers are able to correctly connect them to AI-generated imagery. However, for other features we find stark differences, which allows AI to disguise itself: AI-generated images often incorporate characteristics that consumers typically associate with human-made content: AI generates images with higher aesthetics scores and larger faces — features that consumers generally interpret as signals of human creation. This misalignment between consumer perception and AI capabilities helps explain why some AI-generated content can appear more human-made than it actually is.

Third, our real-world dataset allows us to gain insights into advertisers' adoption of mass-market generative AI tools (see Reisenbichler et al. 2022). Differences exist between advertiser scales and types (i.e., ad arbitrage advertisers), providing important practical implications both for platforms offering AI-powered ad creation tools and advertisers adopting these technologies.

As is inherent in analyzing observational data, this study is not without limitations. Despite the large-scale dataset we employ to construct our quasi-experiments and the various robustness analyses we conduct, our data is still constricted to a single platform, AI model, and time period. However, evaluating large-scale real-world datasets allows us to take a bird's-eye view across advertisers, campaign objectives, and visual features in a manner that is impossible when employing individual A/B tests (Boegershausen et al. 2025). Assessing the impact of AI-generated images on downstream KPIs like conversion rates poses a useful direction for future research. Our data does not cover landing page content and therefore we are unable to isolate and clearly link conversion rate effects to the presence of AI-generated images. Still, for all ads with the campaign objective 'purchases,' we test the effect of AI-generated images on conversion rates (measured as conversions/clicks) and do neither find a significant main effect ($\beta_{AI-generated\,image} = .1649$, $SE = .1103$, $p = .1351$) nor a moderation ($\beta_{AI-generated\,image \times looks-like-AI} = -.0234$, $SE = .1845$, $p = .8991$). This result is consistent with lower-funnel outcomes having lower effect sizes than upper-funnel outcomes (Johnson 2023).[6] Taken together, we find that AI-generated images can boost CTR without sacrificing conversions. Future research should explore this effect with larger sample sizes and additional dependent variables.

Second, we cannot observe the within-day-sequence in which advertisers include different ads within a quasi-experiment (e.g., whether the advertiser created the AI-generated image prior to uploading the human-made image). Such effects may violate Stable Unit Treatment Value Assumption (SUTVA): If advertisers use the GenAI Ad Maker but ultimately choose to submit human-made ads, this would make our control group (human-made images) more similar to our treatment group (AI-generated images), resulting in more conservative estimates. Our robustness check that compares human-made ads that were part of a quasi-experiment with those that were not, helps mitigate such a SUTVA account. To further explore potential learning effects, we compare the first quasi-experiment against its intra-campaign successors but do not find a significant difference between the first and subsequent quasi-experiments

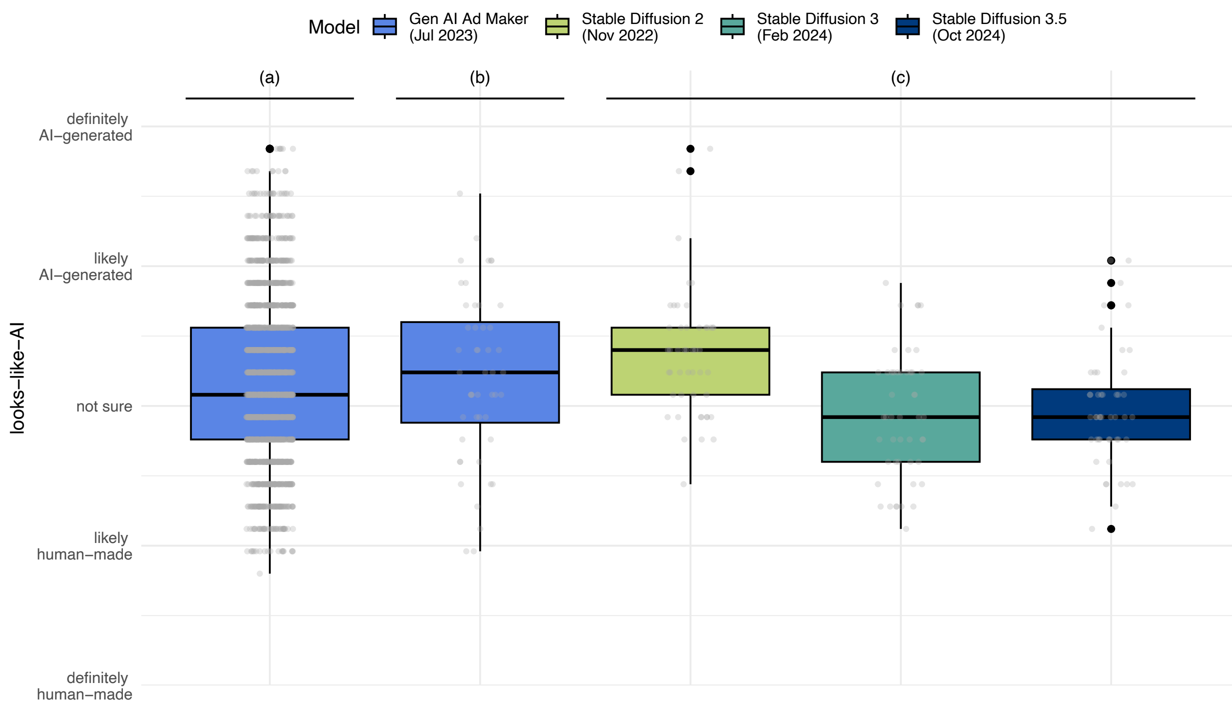
[6] We also tested if other campaign objectives moderated AI-generated images' effectiveness, but found no significant effects.

within a campaign ($\beta_{AI-generated\, image \times first\, quasi-experiment\, in\, campaign} = -.0019$, $SE = .0864$, $p = .9827$). While this analysis provides initial evidence that does not support systematic learning effects within campaigns, the broader question of how advertisers learn to optimize their use of AI tools over time remains an important direction for future research.

Third, our finding on the moderation of perceived artificiality and AI-generated imagery relates to the debate on disclosure of AI content effectiveness of AI-generated imagery (Karpinska-Krakowiak and Eisend 2024). While the platform we worked with follows industry standards and does not disclose whether an ad was generated by humans or AI during the observation period, future research could explore the impact of such disclosures on consumers' reactions to the ad.

Lastly, given the rapid advances of generative AI it is possible that newer tools will generate higher quality and more human-like images (Hartmann et al. 2024), which may affect our results as the perceived artificiality distribution of AI-generated ads will shift towards human-made ratings. This would allow AI-generated images to more effectively disguise themselves. To assess the evolution of text-to-image models, we test if text-to-image models that were introduced post our data periods close the gap of perceived artificiality. Specifically, we compare the degree of artificiality of a stratified sample of 48 ads from our data with newer models. Figure 5 shows that new image generation AI models like Stable Diffusion 3 and 3.5 indeed generate images with lower looks-like-AI scores ($t(47) = -2.3327$, $p = .0240$ and $t(47) = -2.0068$, $p = .0506$, respectively). However, those state-of-the-art models still generate images that cover a broad range of the perceived artificiality spectrum ($\in [0, 1]$), ranging from .28 to .72 and .28 to .76, respectively, and with mean ratings that are insignificantly different from the scale midpoint ($t(47) = 1.0157$, $p = .315$ and $t(47) = .2693$, $p = .7889$, respectively). This persistent variation suggests that some degree of perceived artificiality may persist even as models continue to improve, making our findings about its impact on advertising effectiveness relevant for the foreseeable future.

Figure 5: Range of AI-generated images' perceived artificiality created by newer text-to-image models



Notes. GenAI Ad Maker is based on Stable Diffusion 2 and a proprietary platform-specific prompt specification; model release dates are indicated in parentheses;
(a) all $N = 1{,}751$ images in our quasi-experiments,
(b) a random sample of AI-generated images (stratified by advertiser industry),
(c) images from b) regenerated with newer AI models following Hartmann et al. (2024)'s approach, see Web Appendix Appendix F for an in-depth explanation of this process

Overall, our findings show that there is an art to artificiality: AI performs best if it disguises itself. Based on a large-scale, real-world dataset, we find nuanced patterns in how consumers respond to AI-generated advertising visuals. While AI can match human-level performance in aggregate, its true potential emerges when it creates content that does not look like AI. Our findings open avenues for future research examining the performance of AI in one of the industries that is most ripe for its adoption, advertising. We hope our findings induce future investigations into the complex interplay between artificial intelligence, human perception, and marketing effectiveness in an increasingly AI-enabled advertising landscape.

References

Boegershausen J, Cornil Y, Yi S, Hardisty DJ (2025) On the persistent mischaracterization of Google and Facebook A/B tests: How to conduct and report online platform studies. *International Journal of Research in Marketing* ISSN 01678116, URL http://dx.doi.org/10.1016/j.ijresmar.2024.12.004.

Braun M, de Langhe B, Puntoni S, Schwartz EM (2024) Leveraging Digital Advertising Platforms for Consumer Research. *Journal of Consumer Research* 51(1):119–128, ISSN 0093-5301, URL http://dx.doi.org/10.1093/jcr/ucad058.

Castelo N, Bos MW, Lehmann DR (2019) Task-Dependent Algorithm Aversion. *Journal of Marketing Research* 56(5):809–825, ISSN 0022-2437, URL http://dx.doi.org/10.1177/0022243719851788.

Chui M, Hazan E, Roberts R, Singla A, Smaje K, Sukharevsky A, Yee L, Zemmel R (2023) The economic potential of generative AI: The next productivity frontier URL https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier.

Dentsu (2024) Consumer attitudes toward artificial intelligence (AI) usage by brands in the United States as of March 2024 URL https://www.statista.com/statistics/1462984/consumer-attitudes-ai-brand-usage-united-states/.

Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144(1):114–126, URL http://dx.doi.org/10.1037/xge0000033.

Feeney K (2023) Taboola Launches Generative AI Solution to Revolutionize Ad Creation. URL https://web.archive.org/web/20240606131410/https://blog.taboola.com/generative-ai-ad-creation/.

Hartmann J, Exner Y, Domdey S (2024) The power of generative marketing: Can generative AI create superhuman visual marketing content? *International Journal of Research in Marketing* ISSN 01678116, URL http://dx.doi.org/10.1016/j.ijresmar.2024.09.002.

Horton CB, White MW, Iyengar SS (2023) Bias against AI art can enhance perceptions of human creativity. *Scientific Reports* 13(1):19001, URL http://dx.doi.org/10.1038/s41598-023-45202-3.

Jakesch M, Hancock JT, Naaman M (2023) Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences of the United States of America* 120(11):e2208839120, URL http://dx.doi.org/10.1073/pnas.2208839120.

Johnson GA (2023) Inferno: A guide to field experiments in online display advertising. *Journal of Economics & Management Strategy* 32(3):469–490, ISSN 1058-6407, URL http://dx.doi.org/10.1111/jems.12513.

Karpinska-Krakowiak M, Eisend M (2024) Realistic Portrayals of Untrue Information: The Effects of Deepfaked Ads and Different Types of Disclosures. *Journal of Advertising* 1–11, ISSN 0091-3367, URL http://dx.doi.org/10.1080/00913367.2024.2306415.

Li Y, Xie Y (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, ISSN 0022-2437, URL http://dx.doi.org/10.1177/0022243719881113.

Miller EJ, Steward BA, Witkower Z, Sutherland CAM, Krumhuber EG, Dawel A (2023) AI Hyperrealism: Why AI Faces Are Perceived as More Real Than Human Ones. *Psychological Science* 34(12):1390–1403, URL http://dx.doi.org/10.1177/09567976231207095.

Nightingale SJ, Farid H (2022) AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences of the United States of America* 119(8), URL http://dx.doi.org/10.1073/pnas.2120481119.

Noy S, Zhang W (2023) Experimental evidence on the productivity effects of generative artificial intelligence. *Science (New York, N.Y.)* 381(6654):187–192, URL http://dx.doi.org/10.1126/science.adh2586.

Paivio A, Csapo K (1973) Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology* 5(2):176–206, ISSN 00100285, URL http://dx.doi.org/10.1016/0010-0285(73)90032-7.

Reisenbichler M, Reutterer T, Schweidel DA, Dan D (2022) Frontiers: Supporting Content Marketing with Natural Language Generation. *Marketing Science* 41(3):441–452, ISSN 0732-2399, URL http://dx.doi.org/10.1287/mksc.2022.1354.

Robertson CE, Pröllochs N, Schwarzenegger K, Pärnamets P, van Bavel JJ, Feuerriegel S (2023) Negativity drives online news consumption. *Nature Human Behaviour* 7(5):812–822, URL http://dx.doi.org/10.1038/s41562-023-01538-4.

Taboola (2023) Taboola Beta Tests Integration of Generative AI, Including ChatGPT, to Revolutionize Ad Creation for Advertisers. URL https://web.archive.org/web/20240606131808/https://www.taboola.com/press-release/generativeai.

Talebi H, Milanfar P (2018) NIMA: Neural Image Assessment. *IEEE transactions on image processing* URL http://dx.doi.org/10.1109/TIP.2018.2831899.

Thompson SA (2024) Test Yourself: Which Faces Were Made by A.I.? URL https://web.archive.org/web/20240607121638/https://www.nytimes.com/interactive/2024/01/19/technology/artificial-intelligence-image-generators-faces-quiz.html.

van Heerde HJ, Moorman C, Moreau CP, Palmatier RW (2021) Reality Check: Infusing Ecological Value into Academic Marketing Research. *Journal of Marketing* 85(2):1–13, ISSN 0022-2429, URL http://dx.doi.org/10.1177/0022242921992383.

Zhang S, Lee D, Singh PV, Srinivasan K (2022) What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features. *Management Science* 68(8):5644–5666, ISSN 0025-1909, URL http://dx.doi.org/10.1287/mnsc.2021.4175.