Marketing Science Institute Working Paper Series 2025

Report No. 25-146

AI in Disguise – Quasi-Experimental Analysis of a Large-Scale Deployment of AI-Generated Display Ads

Yannick Exner, Jochen Hartmann, Oded Netzer, Shunyuan Zhang and Ziqian Ding

AI in Disguise – Quasi-Experimental Analysis of a Large-Scale Deployment of AI-Generated Display Ads

Yannick Exner[*] Jochen Hartmann[†] Oded Netzer[‡]
Shunyuan Zhang[§] Ziqian Ding[¶]

Abstract

Recent advancements in generative content creation 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. We collaborate with a leading online ad platform to leverage over 16 billion ad impressions and 116 million clicks. Using this large scale dataset of display ads, we carefully construct a quasi-experimental setup to compare the effectiveness of the resulting 4,633 'sibling ads,' that is, AI-generated and human-made visuals launched by the same advertisers within identical campaign settings at the same time. We find that both for the full dataset and our quasi-experimental sibiling dataset, 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 AI-generated. While AI generates clearer images with larger faces, consumers associate these features with human-made ads. However, in line with consumers' expectations, aesthetics or intense color saturation in ads signals AI generation. 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

[*]TUM School of Management, 80333 Munich. yannick.exner@tum.de. *Corresponding author.*
[†]TUM School of Management, 80333 Munich. jochen.hartmann@tum.de
[‡]Columbia Business School, New York, NY 10027.
[§]Harvard Business School, Boston, MA 02163.
[¶]Tepper School of Business, Pittsburgh, PA 15213. ziqiand@andrew.cmu.edu

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 often more cost effective than human labor, it remains an open empirical question how well AI actually performs in generating ads, and under what conditions.

Initial evidence from controlled lab studies (e.g., Miller et al. 2023, Hartmann et al. 2025) 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., Gelbrich et al. 2025). 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 very challenging, as these tests provide only a narrow view into a specific execution or application context, and depend heavily on researchers' choice of stimuli (Heitmann et al. 2025). Because ads, whether created by AI or by humans, can vary widely in both types and quality, the outcome of any single A/B test may simply reflect the particular ad pairs being compared, rather than provide broader insights into which approach generally performs better and how. Furthermore, from a practical point of view, platforms and advertisers are interested in the learnings from a large scale deployment of AI-generated ads rather than specific anecdotes.

To address this concern, the present paper takes a bird's-eye view, exploiting a large set of actual display ad campaigns generated by real advertisers to assess the effectiveness of AI-generated display ads on consumer behavior. Specifically, we partnered with a leading global ad platform and obtained campaign data covering more than two million daily ad-level observations over a period of 73 days and comprising over 16 billion ad impressions and 116 million clicks. During that period, the platform released the GenAI Ad Maker, a free generative AI-enabled tool that allowed advertisers to opt in to co-create ads using AI alongside their traditional human-made content. The GenAI Ad Maker was introduced platform-wide and the resulting ads span a wide range of product categories, advertisers, and campaign objectives, where consumer responses to AI-generated ads directly impact business profits. Hence, our data and analysis have high ecological validity with economic implications (van Heerde et al. 2021).

A unique feature of our dataset is its nested structure. We utilize the fact that advertisers often employ an experimental mindset by creating multiple ad variations within the same campaign to 'A/B test' the effectiveness of different ad designs (Burtch et al. 2025). Our quasi-experimental setting leverages these naturally occurring 'experiments.' Specifically, we focus on ads 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, yet differing in whether the images were generated by AI or made by humans. This design allows us to employ experiment-specific fixed effects to isolate the effect of AI-generated images on ad effectiveness, while mitigating potential confounders arising from advertisers' decisions to adopt the GenAI Ad Maker in general or use it only for specific campaigns. 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 from an originally generative AI content and thus allows for a cleaner identification of the treatment group.

Overall, our analysis of ad click-through rates (CTR) within this quasi-experimental setting reveals that, on average, AI-generated ads perform comparable to human-made ads, while they can be produced at a fraction of the cost (Hartmann et al. 2025, 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 consumers responding to different AI outputs. Drawing on "algorithm aversion" theory (Dietvorst et al. 2015, Castelo et al. 2019, Zehnle et al. 2025), 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 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). Notably, we find that more than 45% of the AI ads in 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 analysis and find that both at the aggregate and quasi-experimental levels 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. Furthermore, we demonstrate that our looks-like-AI measure is distinct from other visual features such as overall image quality, aesthetics, or realism.

This raises the question of how AI-generated 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. 2025), we find that, compared to human-made images, AI-generated images tend to be, amongst others, more aesthetic, colorful, symmetrical, clearer, and tend to feature larger faces. Although consumers perceive aesthetic ads or ads with vivid colors to be AI-generated, larger faces and clearer images 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, practically, we offer a, one-of-its-kind analysis of the advertisers' early adoption of generative AI tools to enhance their display ad performance. Analyzing AI adoption at scale across advertisers and campaigns provides important practical implications for both platforms and advertisers to enhance their ad effectiveness and efficiency. The platform we collaborated with already adopted some of our insights into their AI generation and interactions with advertisers. Second, substantively, across multiple analyses with varying restrictions on the quasi-experimental settings, we find that ads with AI-generated images achieve human-level CTRs. 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. Finally, we propose a quasi-experimental approach of sibling ads to convert secondary data of ad performance into a quasi-experiment, which allows a more rigorous analysis. The platform we collaborated with is now adopting this method of analysis.

2 Background and data

We collaborated with Taboola, a major US-based global online ad platform, which publishes advertiser-provided ads across publisher websites like MSN, NBC News, or USA TODAY, reaching 500 million daily active users (Feeney 2023). Taboola ads are comparable to Facebook ads in performance.[1]

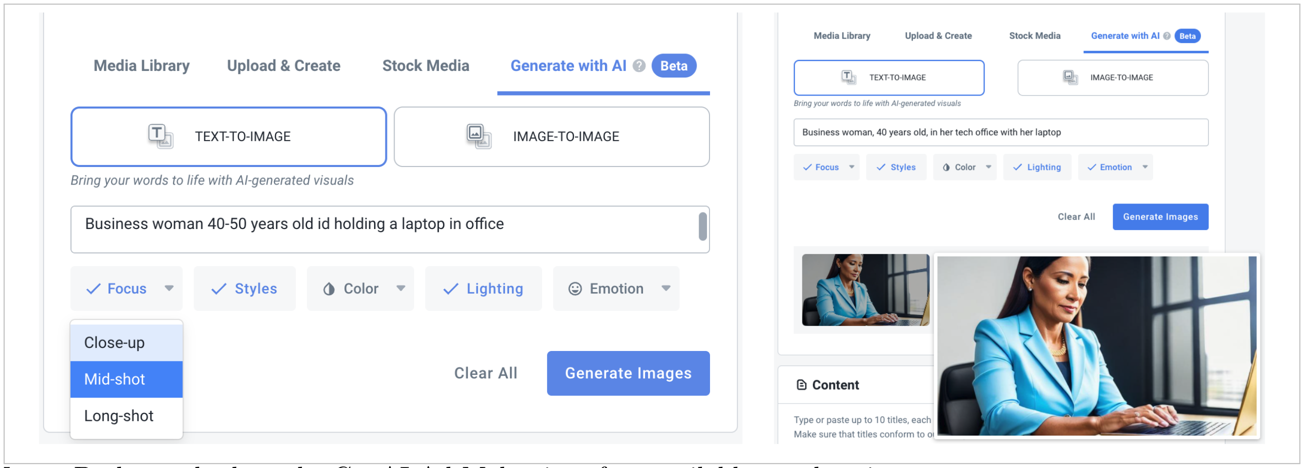
In July 2023, the platform released the GenAI Ad Maker to all English advertisers (Feeney 2023) following a limited beta test that began in February 2023 (Taboola 2023). The launch of the GenAI Ad Maker marked a shift for the platform from its traditional ad publishing role of ad allocation to a more 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.

Typical of display ads, each ad in our data is a unique combination of an image (ad creative) and

[1]See e.g., Hartmann et al. (2025) as well as Gordon et al. (2019) who report a CTR of .57% and an end-to-end conversion rate (i.e., conversions/impressions) of .15% across 15 big field experiments on Facebook totaling over 1.4 billion impressions. Our CTR and end-to-end conversion rate of .71% and .11%, respectively, are in a similar range.

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



Notes: Both panels show the GenAI Ad Maker interface available to advertisers.

a text 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. 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 the textual features of the text caption and whether it was AI-generated.

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 full 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%.

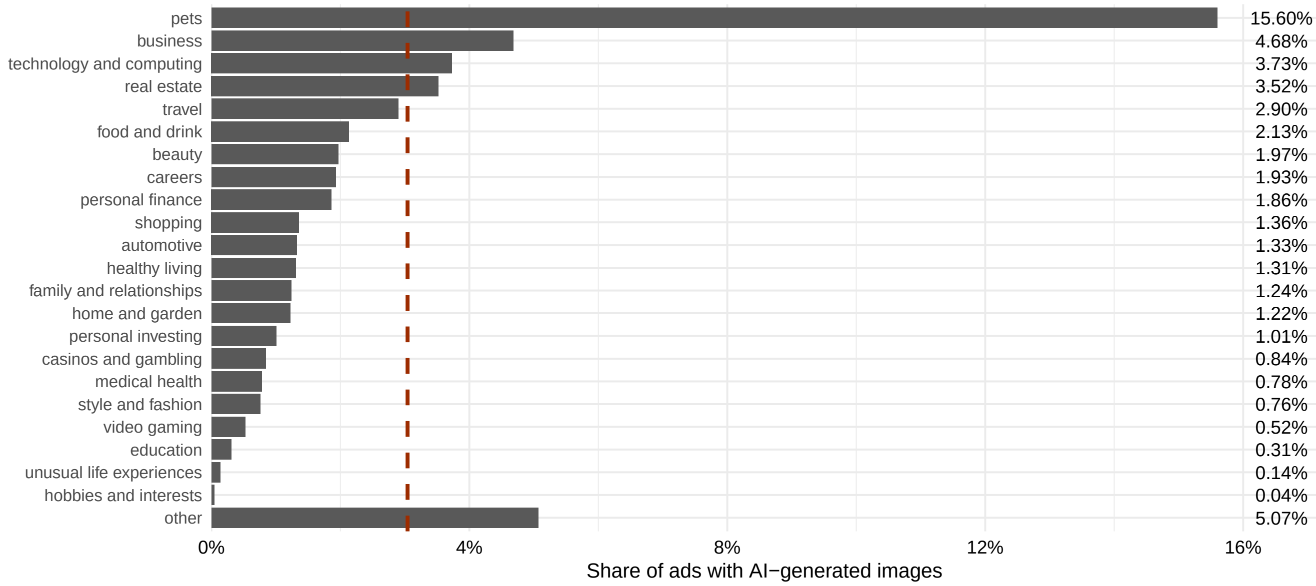
2.1 How advertisers use the GenAI Ad Maker

Before assessing the effectiveness of AI-generated ads on consumers, it is important to gain insights into the supply side of the AI-enabled advertisement market, that is, to understand which and how advertisers use the AI tool. For platforms like Taboola, adoption patterns among advertisers determine whether a new tool such as the GenAI Ad Maker meaningfully changes the advertising ecosystem. While our data are limited to only one platform with one AI tool and time period, that is, the first 34 days after the GenAI Ad Maker's go-live, having a bird's-eye view across thousands of advertisers on the platform can be informative about the pattern of AI adoption in industry beyond the narrow view of most existing academic works (e.g., Hartmann et al. 2025) and industry reports (e.g., Chui et al. 2023). For example, if adoptions vary across verticals, these differential patterns can provide valuable guidance to Taboola and other platforms on how to position, promote, and support its AI tools, while informing the platform's long-term AI strategy.

In our data, we observe which advertisers publish AI-generated ads and for what campaigns.[2] 6.25% of the advertisers used the GenAI Ad Maker at least once by the end of our observation period. Across all ads, the advertisers used the AI tool to generate 3.04% of the images after the GenAI Ad Maker went live. We observe heterogeneity in the adoption of the GenAI Ad Maker across advertiser industries. Figure 2 presents the tool adoption rates for AI-generated marketing imagery within the 34 days following the tool's go-live. Advertisers in the categories *pets*, *business*, *technology and computing*, and *real estate* adopt the GenAI Ad Maker more frequently to generate ad visuals compared to the remaining industries.

[2]Our dataset 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.

Figure 2: Adoption of the GenAI Ad Maker's image generation feature across advertiser industries



Notes: The vertical dashed line at 3.04% indicates the mean adoption rate across all industries post go-live. Industries with less than 1,000 ads are clustered as *other*. Non-adopting industries are omitted.

Looking beyond the industry verticals, to further understand which advertisers adopt the GenAI Ad Maker, we regress different adoption metrics on a set of advertiser characteristics, observed during the 39 days prior to the GenAI Ad Maker's go-live, including ad count, CTR, impressions, and spend in USD. Additionally, we include in the regression the campaign objective, industry, and if an advertiser is an arbitrage advertiser[3]. Table 1 Model 1 through 3 present the results of regression models examining which advertiser characteristics are associated with adoption of GenAI Ad Maker. We use three different dependent variables to capture adoption: (1) whether an advertiser adopts the tool at least once during the data period, (2) the within-advertiser share of ads that are AI-generated, and (3) the within-advertiser share of campaigns that include at least one AI-generated ad.

We do not find any statistically significant relationship between advertisers' pre go-live characteristics on their adoption of the GenAI Ad Maker's image generation feature (see Table 1 Models 1 through 3). Specifically, it does not seem that more active advertisers (higher spending) or more successful advertisers (higher CTR) are associated with higher AI adoption. Turning to campaign level characteristics, we find that advertisers are more likely to adopt GenAI for conversion-oriented

[3]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.

Table 1: Relationship between advertisers' and campaign's characteristics and GenAI Ad Maker adoption

Dependent Variables:	Advertiser adopts	Share of AI ads	Share of campaigns with AI ads
Model:	(1)	(2)	(3)
Variables			
Constant	-19.34***	-19.99***	-19.19***
	(1.115)	(1.073)	(1.127)
Pre go-live characteristics			
Ad count per advertiser (log)	-.0369	.0661	.1047
	(.1576)	(.2057)	(.1898)
Prior CTR	-1.718	-9.691	-15.79
	(6.371)	(10.19)	(12.88)
Impressions (log)	.1861	.0813	.0072
	(.1109)	(.1181)	(.1174)
Spend (log)	-.1912	-.1407	-.0767
	(.1024)	(.1137)	(.1108)
Post go-live characteristics			
Campaign objective: Brand awareness	.9202	1.587	1.434
	(.8407)	(.8711)	(.7952)
Campaign objective: Mobile app installs	1.320	1.501*	1.401*
	(.7873)	(.7336)	(.6932)
Campaign objective: Page views	.4980	.8219	.6635
	(.3267)	(.4229)	(.3673)
Campaign objective: Purchases	.7607**	.9084*	.7762*
	(.2890)	(.3698)	(.3637)
Controls			
Arbitrage advertiser	Yes	Yes	Yes
Industry	Yes	Yes	Yes
Fit statistics			
Observations	1,571	1,571	1,571
Squared Correlation	.09988	.16518	.14076
Pseudo R^2	.12883	.33214	.31392
BIC	1,230.7	782.64	864.77

*Signif. Codes: ***: .001, **: .01, *: .05*

Notes: We exclude all advertisers that a) participated in the GenAI Ad Maker's beta phase ($N = 36$), b) have missing industry information ($N = 79$), and were inactive either c) prior ($N =$ 1,850) or d) after the GenAI Ad Maker's go live ($N =$ 3,538). We obtain directionally consistent results when estimating OLS regressions for all three models.

campaigns. Specifically, advertisers with a higher share of campaigns aimed at purchases or mobile app installs are significantly more likely to adopt the GenAI Ad Maker's image generation feature.

From a managerial perspective, the results imply that advertising platforms like Taboola can accelerate adoption by targeting AI tool promotion and support toward performance-oriented advertisers, who appear most receptive to experimenting with new AI generative technologies. These adoption patterns also highlight the importance of accounting for campaign objectives in our quasi-experimental design (which we detail in the next subsection), since advertisers with different goals may systematically differ in their likelihood of adopting AI and possibly in how their audiences respond to AI-generated ads.

2.2 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).

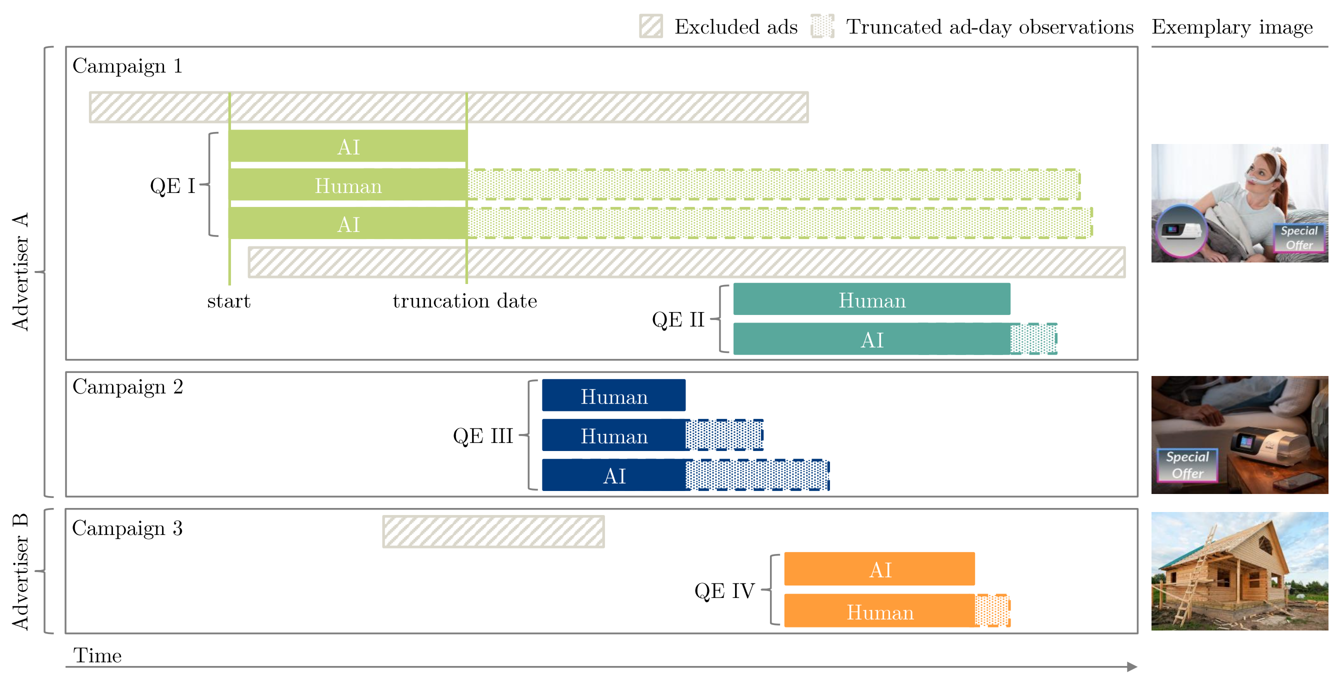
(iv) Advertisers may keep more effective ads in the field for a longer period of time, which may result in more ad-day observations for effective ads versus less effective ads.

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

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

To address these challenges, we construct a quasi-experimental setting in a three-step approach (see Figure 3 for a visual schema of our approach). Typically, advertisers run multiple ads in a single campaign to test different ad variations with an experimental mindset (Burtch et al. 2024). 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 3: 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) in our quasi-experimental analysis we only assess campaigns that include both ads with AI-generated and human-made images.[4] To address challenge (iii) and (iv): First, we require all ads to be created on the same day and thus be part of the same quasi-experiment within a campaign. This design captures the experimental mindset of advertisers who launch multiple ad creatives at the same time, and it allows us to include a quasi-experiment fixed effect to explore within campaign and within advertiser variation. Thus, we mitigate potential concerns regarding endogenous decisions by the advertisers to adopt the GenAI Ad Maker in general or for specific campaigns. Additionally, by restricting our analysis to ads created on the same day, we eliminate any confounding effects from staggered ad roll-outs within a campaign,

[4]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 E.10).

ensuring that temporal factors do not bias our treatment effect estimates. For example, campaign 1 in Figure 3 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). Second, we truncate all ads once the first ad in a quasi-experiment stops to ensure comparable runtimes and address a potential 'survival bias' as defined in challenge (iv) (see dotted bars to the right of the 'truncation date' in Figure 3 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.). Taken together, 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 performs differently than 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. In addition, we conduct a robustness check and exclude human-made images that are detected as AI-generated by a state-of-the-art AI detector based on masked spectral learning. That analysis led to similar results as the main analysis (see Web Appendix E.9). Lastly, 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 (Coffee 2025, Boegershausen et al. 2025, Braun et al. 2024, Burtch et al. 2025). 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, we believe that our results are subject to less divergent delivery compared to data from other ad platforms. First, due to its relatively limited first-party data on user characteristics, the Taboola platform algorithmic targeting is not able to target on user-level demographics and granular characteristics to the same extent as larger platforms such as Meta could (Burtch et al. 2025). Second, our industry partner confirmed that there was no indication to the targeting algorithm whether an ad is AI-generated or whether the advertiser used the GenAI Ad Maker. Third, we find that the total exposure of ads in terms of daily impressions (log transformed) did not differ between the experimental groups, mitigating our concern about algorithmic differences driving the observed effects ($\beta_{AI-generated\ image} = -.0631$, $SE = .1153$, $p = .5842$). Our results are robust to controlling for the degree of imbalance in impressions across conditions (see Web Appendix Table 9). Lastly, as we demonstrate in the robustness checks section, we empirically test for the risk of divergent delivery by repeating the analysis focusing only on the first few days of a quasi-experiment, before the platform's targeting algorithm has had sufficient time to gather information to optimize ad delivery based on performance signals (Boegershausen et al. 2025). Our results during the first few days of campaign launch are similar to our main results.

Thus, in addition to analyzing the larger dataset of over 2 million daily observations, our more controlled quasi-experimental setting, results in a dataset that 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 data 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 A for further details.

3 Consumer reaction to AI-generated ads

Using the constructed quasi-experimental 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 outcome variable is CTR, measured as clicks divided by impressions, a commonly used measure to assess ad performance (e.g., 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 an ad's 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 cluster $j = 1, ..., J$. A cluster refers to the grouping within which we compare AI-generated to human-made ads (i.e., either within an advertiser, campaign, or quasi-experiment) depending on how restrictive we construct the quasi-experimental setup. We analyze the data at the level of an advertiser, a campaign and finally the most restrictive quasi-experiment as discussed above. $t = 1, ...T$ reflects the calendar date an observation was recorded on. Accordingly, θ_{ijt} refers to the CTR of ad i in cluster 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 cluster 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[5]:

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

[5]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.

where x_{ij} is the effect-coded variable indicating if ad i in cluster j uses an AI-generated image ($x_{ij} := +1$ for AI-generated and $x_{ij} := -1$ for human-made content). Our primary focus is on the coefficient β_1. This coefficient represents the change in log odds of ad i's CTR on day t when using an AI-generated image relative to a human uploaded image. X_{ijt} is a matrix of control variables including an indicator if the ad's caption is AI-generated with the GenAI Ad Maker as well as a set of verbal features in the ads caption and description (e.g., word count or authenticity), and visual features of an ad's image (e.g., contrast or aesthetics). Web Appendix B and 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 cluster j (advertiser fixed effect in Model 1, campaign fixed effect in Model 2, and quasi-experiment fixed effect in Model 3) and calendar date t, respectively. All continuous independent variables are z-standardized for increased interpretability.

Model 1 of Table 2 presents the results from estimating Equation 2 on the full dataset with an advertiser fixed effect prior to any filtering and quasi-experimental settings. As can be seen, AI-generated images seem to perform significantly better than human-made images in terms of their CTR ($\beta_{AI-generated\,image} = .1272$, $SE = .0524$, $p = .0153$). Next, to address challenges (i) and (ii), we estimate Model 2 and restrict comparisons to ads within a campaign, thereby controlling for advertiser- and campaign- heterogeneity using a campaign fixed effect. The results indicate a significant superior performance of AI-generated versus human-made images, though with a smaller effect size ($\beta_{AI-generated\,image} = .0446$, $SE = .0227$, $p = .0498$). Finally, in Model 3 we address concerns about staggered ad launches and differences in ad run-times using our full quasi-experimental design. The results suggest that, in this most stringent setting, AI-generated images maintain on par performance versus human-made ads when evaluated under tightly controlled conditions ($\beta_{AI-generated\,image} = -.0525$, $SE = .0329$, $p = .1103$).

Table 2: Performance of AI-generated images

Dependent Variable:	CTR		
	All observations	Campaigns with AI & human ads	Right truncation
Model:	(1)	(2)	(3)
Variables			
AI-generated image	.1272*	.0446*	-.0525
	(.0524)	(.0227)	(.0329)
Controls			
Caption	Yes	Yes	Yes
Description	Yes	Yes	Yes
Image	Yes	Yes	Yes
Fixed effects			
Calendar date	Yes	Yes	Yes
Advertiser	Yes		
Campaign		Yes	
Quasi-experiment			Yes
Fit statistics			
Observations	2,218,322	40,044	29,592
Squared Correlation	.03939	.06964	.19608
Pseudo R^2	.95634	.97985	.98335
BIC	60,720,520.6	835,352.1	516,040.9

*Signif. Codes: ***: .001, **: .01, *: .05*

Notes: AI-generated image is effect-coded. Model 1, 2, and 3 report clustered standard errors at the advertiser, campaign, and quasi-experiment level, respectively (shown in parentheses). Observation count in Model 3 differs slightly from 29,631 reported in Web Appendix A due to lack of variation in the DV for 39 (0.13%) observations.

First, these results suggest that AI-generated images in ads, while likely much more cost-efficient to produce, elicit consumer responses that are insignificantly different from human-made images in ads. This highlights the potential of generative AI technologies to democratize the ad creation and broaden access to effective advertising without sacrificing performance. Additionally, looking at the change in the magnitude of the AI-generated image coefficient across models, these results highlight the potential self-selection in advertisers' use of GenAI and the need to conduct a quasi-experimental analysis.

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 these questions, 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: whether human observers believe it was generated by AI or made by humans (i.e., perceived artificiality[6]).

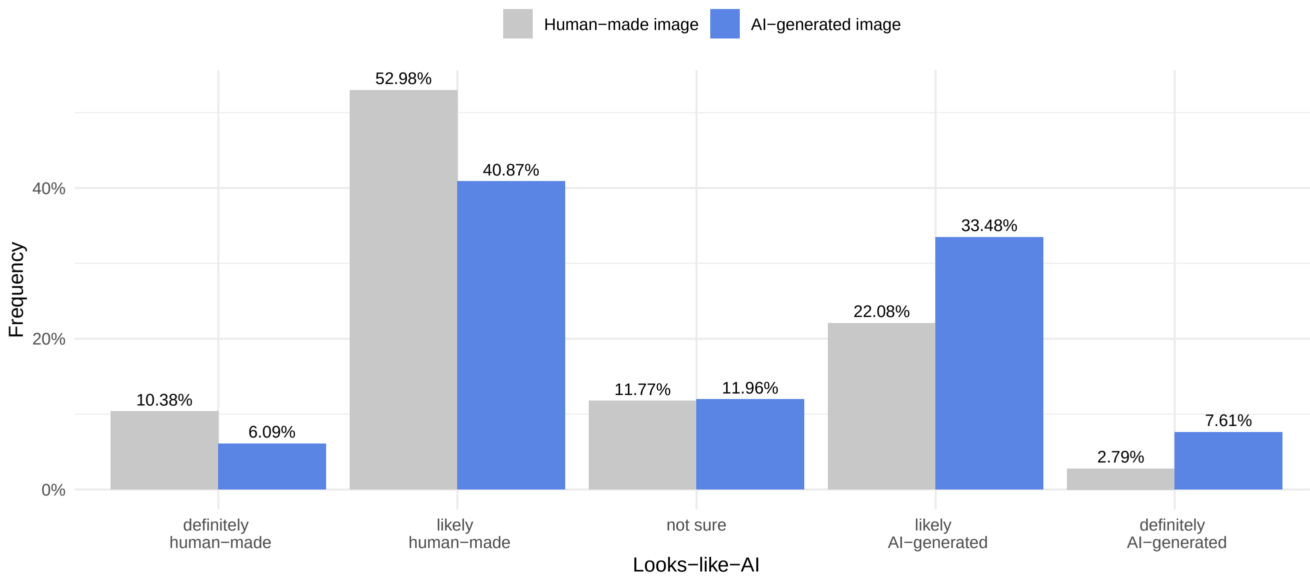
For each of the unique images in our quasi-experiments from Table 2 Model 3 (N = 1,751, with N = 460 AI-generated and N = 1,291 human-made images), we ask five MTurkers to rate it on a 5-point Likert scale: 'Is this image human-made OR AI-generated?' (adopted from Jakesch et al. 2023). We normalize the ratings on perceived artificiality $\in [0, 1]$ (see Web Appendix D for further details).

Figure 4 displays the distribution of perceived artificiality. While AI-generated images (M = .4891, SD = .2842) have a significantly higher artificiality score than human-made ones (M = .3848, SD = .2581; $t(745) = 6.9235$, $p < .001$), humans struggle to clearly 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 5 illustrates exemplary images from our data, juxtaposing a 2 $\times$ 2 comparison of ad content by their true source (AI-generated versus human-made) as well as by their perceived artificiality (human-made versus AI-generated), along with the average CTRs observed across all ads in that group. This model-free evidence suggests that: AI-generated images that disguise their origin (i.e., are rated as 'definitely human-made,' 'likely human-made,' or 'not sure') have a mean CTR of .79% and significantly outperform the mean .62% CTR of AI-generated images

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

Figure 4: Perceptual human ratings on ads' perceived artificiality



Notes: $N = 1{,}751$, thereof $N = 1{,}291$ human-made and $N = 460$ AI-generated images.

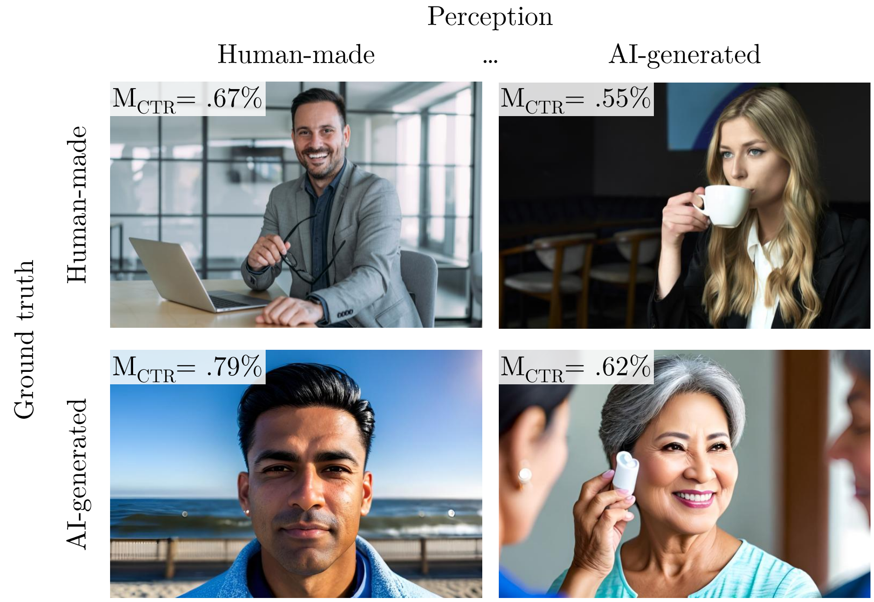
that do not disguise their origin (i.e., are perceived as 'likely' or 'definitely' AI-generated), as well as human-generated ads that look or do not look like AI (.55% and .67%, respectively). These descriptive results suggest that what matters for ad performance is less about the actual source of an image - AI or human-made - and instead more on how consumers perceive it.

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' likelihood of clicking on an ad vary based on their perception of who created the ad (i.e., human or machine) and based on the degree of an ad's perceived artificiality. Specifically, we replicate the regressions from Table 2 and control for looks-like-AI[7] and its interaction with the effect-coded indicator for an AI-generated image. Our interest is the coefficient of interaction term $\beta_{AI-generated\ image \times looks-like-AI}$: a positive (negative) interaction coefficient would suggests that AI-generated ad images perform better when the perceived artificiality is high (low).

Models 1 to 3 in Table 3 present the results for our effect-coded indicator 'AI-generated image' and our mean-centered measure 'looks-like-AI'.

[7] We predict looks-like-AI for the remaining $N = 49{,}141$ images in our full data that are not part of the quasi-experimental setup in Model 3 for which we collected the $N = 1{,}751$ looks-like-AI ratings. The XGBRegressor model predicts looks-like-AI $\in [0, 1]$ with a hold-out MSE of .0662 and correlation to actual values of $\rho = .6843$, $p < .001$. We obtain consistent results when using the actual values from Model 3 (see Web Appendix Table 12).

Figure 5: 2 × 2 of ad generation sources and ad perception with exemplary images, and average CTR per cell



Notes: M_{CTR} are average CTRs for all ads in that cell in our dataset, not just for the sample ad presented.

The patterns shown in Table 3 are similar across the models. Consistent with Table 2 the main effect of whether the ad was generated by AI is positive and significant in Models 1 and 2 and insignificant in Model 3. Thus, on average, consumers do neither discriminate nor favor such images in terms of CTR, helping advertisers to leverage efficiency increases in content-generation without impacting downstream advertisement effectiveness.

However, consistent with the literature on AI aversion, we find that consumers' perception of content as AI-generated has a strong negative effect on CTR ($\beta_{looks-like-AI} = -1.969, SE = .3633, p < .001$ in Model 3): consumers seem to have a negative predisposition to what they perceive as AI-like – irrespective of whether an image is actually AI-generated or human-made.

Importantly, the interaction term is negative and significant across all three models, indicating that AI-generated ads are especially 'penalized' when they appear AI-like ($\beta_{AI-generated\ image \times looks-like-AI} = -.9259$, $SE = .2545$, $p < .001$; see Table 3 Model 3 and Web Appendix Figure 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 disposition to perceived artificiality in AI-generated versus human-made content (Dietvorst et al. 2015).

Furthermore, the combined effect (i.e., main effect plus interaction) shows that AI-generated images achieve a positive and significant net advantage whenever they are not perceived as AI.

Table 3: CTR performance of AI-generated images depending on consumers' perceived artificiality (looks-like-AI)

Dependent Variable:		CTR	
	All observations	Campaigns with AI & human ads	Right truncation
Model:	(1)	(2)	(3)
Variables			
AI-generated image	.1250**	.0665**	-.0023
	(.0408)	(.0238)	(.0304)
looks-like-AI	-1.059***	-1.303***	-1.969***
	(.2209)	(.3330)	(.3633)
AI-generated image × looks-like-AI	-1.151***	-.5307*	-.9259***
	(.2883)	(.2314)	(.2454)
Controls			
Caption	Yes	Yes	Yes
Description	Yes	Yes	Yes
Image	Yes	Yes	Yes
Fixed effects			
Calendar date	Yes	Yes	Yes
Advertiser	Yes		
Campaign		Yes	
Quasi-experiment			Yes
Fit statistics			
Observations	2,218,322	40,044	29,592
Squared Correlation	.03938	.06921	.19639
Pseudo R^2	.95636	.98006	.98365
BIC	60,693,510.8	827,003.9	507,032.3

*Signif. Codes: ***: .001, **: .01, *: .05*

Notes: AI-generated image is effect-coded. Looks-like-AI is scaled $\in [0, 1]$ and mean-centered. Models 1, 2, and 3 report clustered standard errors at the advertiser, campaign, and quasi-experiment level, respectively (shown in parentheses).

While Models 1 and 2 are directionally consistent, going forward, we focus on Model 3, as it is based on our most stringent quasi-experimental setting, addressing our challenges as outlined in the previous section and providing the cleanest setting for interpreting consumer responses.

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 E. Specifically, we obtain consistent results when we:

(1) exclude short-running ads with runtimes smaller than one, two, three, or four days, where advertisers may have simply 'played around' with the GenAI Ad Maker as well as when

ignoring the truncation requirement from challenge (iv).

(2) exclude ads with less than 1,000 impressions or less than < 100 clicks.

(3) exclude ads with CTRs above one, two, or three standard deviations above the mean.

(4) focus only on the first one to seven 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) account for potential imbalances in daily impressions across different ads in a quasi-experiment.

(6) exclude pre-go-live ads and advertisers who participated in the GenAI Ad Maker's beta phase.

(7) use the mean-based looks-like-AI ratings on the quasi-experimental level instead of the median-based reference.

(8) use the actual looks-like-AI ratings on the quasi-experimental level (Model 3) instead of the predicted values.

(9) exclude human-made images that are detected as AI-generated, based on state-of-the-art AI detectors.

Lastly, 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, suggesting that advertisers do not leverage AI-generated ads particularly when their human-made ads are less good.

5 Disguising AI: Visual features and consumer perception

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

Having established the relationship between AI-generated images and ad performance, and the role of perceived artificiality, we turn our attention to understanding what makes an image look like AI. Understanding these deep-seated perceptual mechanisms is crucial (Huang and Katona 2025), as Table 3 highlights that advertisers can benefit from higher ad performance if their ad visuals exhibit lower levels of perceived artificiality and that AI-generated ads are particularly effective if its image does not look like AI. While generative AI models 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, 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. 2025). Web Appendix C describes these 23 fundamental features in more detail.

We regress the looks-like-AI score of each ad image present in our quasi-experiments on their visual features to explore the determinants of looks-like-AI and assess how they drive consumers' perceived artificiality. Web Appendix Table 15 Model 1 presents the corresponding results where continuous features are z-transformed for ease of interpretation.

One might wonder, if perceived artificiality is simply a proxy relating to how much consumers like or dislike an image. However, our findings indicate that consumers' perception of an image's perceived artificiality is not significantly related to its perceived quality (quality ratings were collected from consumers via MTurk; $\beta_{quality} = .0019$, $SE = .0075$, $p = .0804$).

Further, we find that consumers perceive images with higher feature values for aesthetics and color saturation as rather AI-generated. In contrast, people perceive images with high realism, clearness, warmth, facial area, and a technology theme as rather human-made, being associated with lower perceived artificiality (realism ratings were collected from consumers via MTurk). The association of facial presence with looks-like-AI is in line with Miller et al. (2023) who find that humans can perceive AI-generated faces as "more real than human ones."

Next, we explore whether consumers' lay beliefs what makes an image look like AI are consistent with visual features that are actually more likely to appear in AI-generated ad images.

5.2 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 an OLS 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 Web Appendix Table 15, Model 2.[8]

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 clear, more colorful, feature more contrast, have higher levels of Kurtosis, are more symmetrical, have more depth of field, inhibit higher entropy (i.e., a proxy for the amount of information in the image), and are more complex. Additionally, AI-generated images are significantly more likely to be food- or sports-themed. Our findings align with Wedel and Pieters (2015)'s finding that "color composition of the central object plays a key role" in consumers' perception of advertisements (see e.g., Web Appendix Figure 2 for images with different color saturation). Interestingly, AI-generated images also create more and larger faces. Compared to human-designed ad imagery, they are significantly more likely to display (large) faces.

Our findings also suggest that 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. 2025). We further find that AI-generated images are less blurry, warm, and diagonally dominant (i.e., diagonal alignment) versus human-made images.

[8] We obtain directionally consistent results when estimating a logistic regression instead of an OLS. Web Appendix Figure 1 presents model-free evidence of the prevalence of different visual features in AI and human-made images.

5.3 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. 2025), our analysis reveals both alignment and misalignment between perception and reality for visual features in AI-generated images.

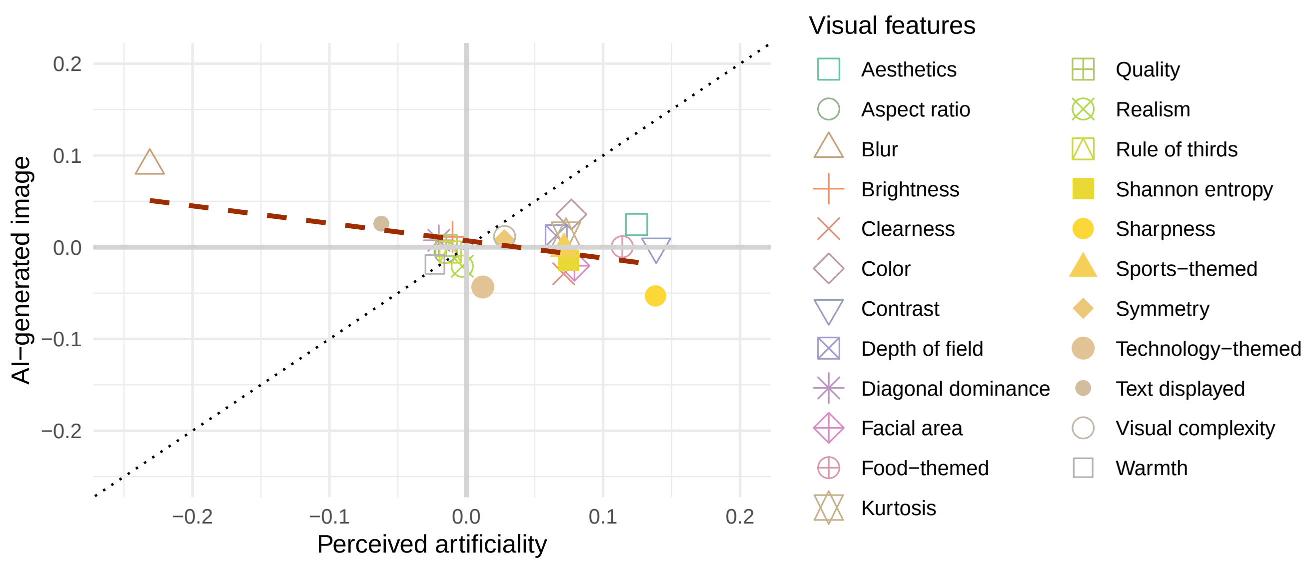
In some cases, consumer intuition proves in line with AI behavior. For example, the more aesthetic or colorful an image is, the more likely consumers are to identify it as AI-generated. Likewise, the less warm an image is, the less likely consumers are to identify it as AI-generated, matching actual characteristics of AI-generated images in both instances. However, AI effectively disguises itself through other features. Specifically, AI generates images with higher clearness and larger faces – features that consumers typically associate with human-made content. These findings demonstrate how AI can leverage misaligned perceptions, particularly through larger face size and higher clearness, to appear more human-made than it actually is.

Across all visual features, we find a strong and significant negative 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 Web Appendix Table 15 Models 1 and 2 is $\rho = -.5323$, $p = .0089$), suggesting that the AI-generated images successfully disguise their origin, by deviating from consumer perceptions of what looks like AI. Figure 6 displays these regression coefficients which are dispersed from the identity line ($f(x) = x$). For advertisers, an effective use of generative AI is perhaps less about whether content is AI-generated or human-made, and more about selecting or guiding AI outputs to avoid the perceptual markers of artificiality.

6 General discussion

In collaboration with a leading ad platform, and using large-scale, real-world data we explore the deployment of a GenAI-enabled tool in online display advertising. We construct a carefully constructed quasi-experimental setup to assess the performance of AI-generated marketing images in display ads. We find nuanced patterns in consumers' response to AI-generated advertising visuals that are both present in the full data as well as in the resulting quasi-experimental setup: On average, AI-generated ads perform insignificantly different from human-made ads. However,

Figure 6: Comparison of feature contribution to perceived artificiality versus actual AI generation



Notes: Regression coefficients for perceived artificiality and AI-generated image from Web Appendix Table 15 Models 1 and 2, respectively. The dashed red line represents a fitted regression line. The dotted gray 45° line is the identity line ($f(x) = x$) representing perfect correlation.

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. AI performs best if it disguises itself.

Our research has implications for both theory and practice. First, we demonstrate real-world empirical evidence on the effectiveness of AI-generated display ads. Given the tremendous efficiency gains[9] that come along with AI-generated content (e.g., Reisenbichler et al. 2022, Noy and Zhang 2023, Hartmann et al. 2025), 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 when distributing these images in the field. Irrespective of an image's origin as AI-generated or human-made, we find that higher levels of perceived artificiality seem to lower an ad's performance. Advertisers can benefit from tailoring marketing visuals to ensure low levels of perceived artificiality. Further, we contribute to the understanding of how AI-generated content in particular can outperform human-made images in effectiveness. Importantly, we do not only show that AI works, but also when it does *not* work, identifying ads' perceived artificiality as an

[9]The GenAI Ad Maker comes at zero cost to the advertisers. Discussion with the collaborating ad platform suggest a cost of image generation incurred to platform of no more than 2 cents per GenAI Ad Maker request (one request includes four image candidates which each cost approx. USD .005 at 768x512 pixels).

important boundary condition: AI-generated images achieve superhuman CTR levels only when they do not appear to be AI-generated. Correspondingly, our findings have relevant managerial implications: Advertisers as well as platforms could apply the looks-like-AI measure to screen AI-generated images before deploying them. They could further utilize our findings to either prompt engineer or fine-tune AI models to generate images with lower perceived artificiality. Specifically, the company we collaborated with has started to implement these insights into their ad creatives.

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 aesthetics, color, and warmth, consumers are able to correctly link 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 clearness 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 early insights into advertisers' actual adoption of mass-market generative AI tools (see Reisenbichler et al. 2022). Differences exist across advertiser industries and campaign-specific marketing objectives, providing important practical implications for both advertisers adopting these technologies and platforms offering AI-powered ad creation tools.

Fourth, our method of creating quasi-experimental data from observational advertising data using a 'sibling' design can be useful for platforms to investigate ad effects with high ecological validity. The company we collaborated with found the designs we proposed useful for their analysis of ads beyond assessing adoption of AI.

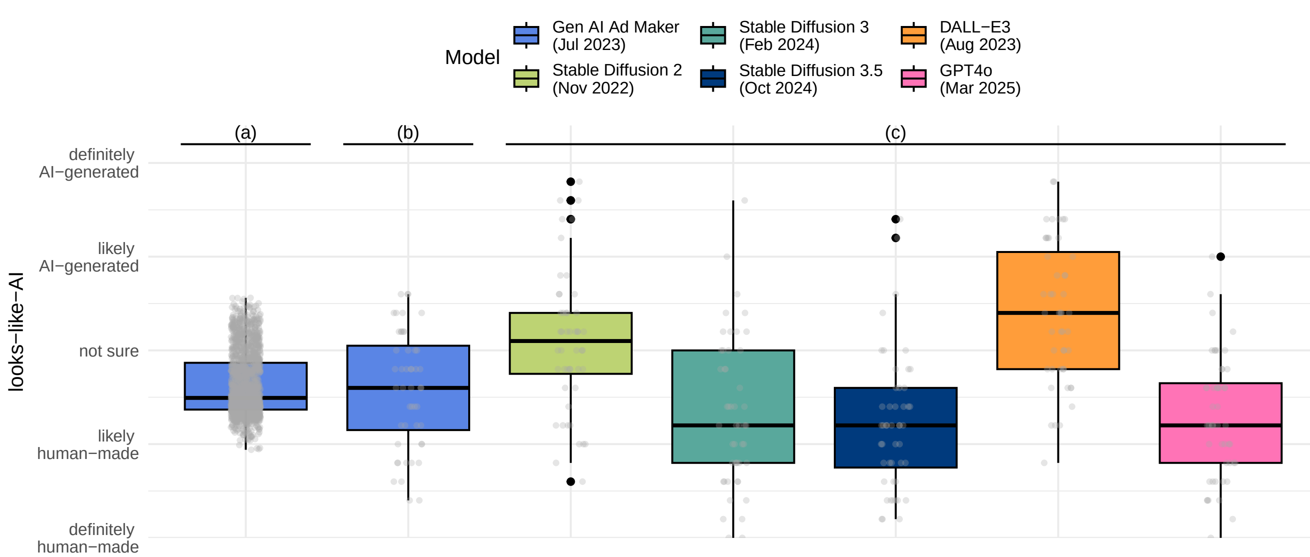
As is inherent in analyzing observational data, this study is not without limitations. Despite the large-scale dataset we employ to construct our quasi-experiments and the various robustness analyses we conduct, our data is still constricted to a single platform, AI model, and time period.

However, evaluating real-world data 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 outcomes 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. Nevertheless, we test the effect of AI-generated images on conversion rates (measured as conversions/clicks) for our quasi-experimental data and do neither find a significant main effect ($\beta_{AI-generated\,image} = -.0407$, $SE = .0339$, $p = .2302$) nor a looks-like-AI or a moderation effect ($\beta_{looks-like-AI} = -.3514$, $SE = .4434$, $p = .4281$ and $\beta_{AI-generated\,image \times looks-like-AI} = .1783$, $SE = .3256$, $p = .5840$, respectively). This result is consistent with lower-funnel outcomes having lower effect sizes than upper-funnel outcomes (Johnson 2023). 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 more nuanced measures of conversions.

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 within a campaign ($\beta_{AI-generated\,image \times first\,quasi-experiment\,in\,campaign} = .0040$, $SE = .0625$, $p = .9486$). 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 given that most AI-based workflows follow a human-in-the-loop concept (Reisenbichler et al. 2022). Additionally, our data does not allow us to understand how advertisers screen and select images within the GenAI Ad Maker. Future research could assess advertisers' screening and selection process in common human-in-the-loop settings.

Looking into the foreseeable future, two forces shaping the future of AI-enabled advertising may have an impact on our findings. First, given the rapid advances of generative AI (Zehnle et al. 2025) newer tools will likely generate higher quality and more human-like images (Hartmann et al. 2025). To assess such advancements, we test if image AI models that were introduced after our data period close the gap of perceived artificiality. Specifically, we compare the degree of artificiality of an industry-stratified sample of 48 ads from our data with newer models. Figure 7 shows that new AI image generation models like Stable Diffusion 3.5 or ChatGPT 4o generate images with directionally lower looks-like-AI scores ($t(47) = -.9946$, $p = .3250$ and $t(47) = -.9142$, $p = .3653$, respectively). However, these images still cover a broad range of the perceived artificiality spectrum ($\in [0, 1]$), ranging from .30 to .58 and .28 to .57, respectively. This enduring 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 7: Range of AI-generated images' perceived artificiality created by newer GenAI 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. (2025)'s approach; see Web Appendix G for an in-depth explanation of this process.

Second, the regulatory environment is shifting. While the platform in our study did follow industry standards and did not disclose whether an ad was created with the GenAI Ad Maker during our observation period, several jurisdictions now legally mandate AI-generated content to be disclosed (see e.g., California State Legislature 2025), whereas others do not. This de jure heterogeneity in disclosure policies as well as de facto user behavior (e.g., so called "AI slop" Mahdawi 2025) indicates that our findings maintain relevant given the continued presence of undisclosed AI-generated content on the web. Future research should add to the initial evidence on such disclosure effects (e.g., Karpinska-Krakowiak and Eisend 2024) and examine how mandated disclosure interacts with perceived artificiality in shaping consumer reactions to AI-generated ads.

Overall, our research demonstrates that there is an art to artificiality: AI performs best if it disguises itself. We find nuanced patterns in how consumers respond to AI-generated advertising visuals in carefully curated quasi-experimental settings that align with what we find in the full data. 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. As generative models improve and perceptions shift, the boundary between AI-generated and human-made ads may start or continue to blur, making the question of how content looks to consumers even more central to both academic research and advertising practice.

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.

Burtch G, Lee D, Chen Z (2024) The consequences of generative AI for online knowledge communities. *Scientific Reports* 14(1):10413, URL http://dx.doi.org/10.1038/s41598-024-61221-0.

Burtch G, Moakler R, Gordon BR, Zhang P, Hill S (2025) Characterizing and Minimizing Divergent Delivery in Meta Advertising Experiments. *arXiv* URL http://dx.doi.org/10.48550/arXiv.2508.21251.

California State Legislature (2025) California AI Transparency Act. California Assembly Bill 853 Chapter 674, URL www.leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202520260AB853.

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 www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier.

Coffee P (2025) AI Will Soon Dominate Ad Buying, Whether Marketers Like It or Not.

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 blog.taboola.com/generative-ai-ad-creation/.

Gelbrich K, Roschk H, Miederer S, Kerath A (2025) Automated versus human agents: A meta-analysis of customer responses to robots, chatbots, and algorithms and their contingencies. *Journal of Marketing* 0(0):00222429251344139, URL http://dx.doi.org/10.1177/00222429251344139.

Gordon BR, Zettelmeyer F, Bhargava N, Chapsky D (2019) A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. *Marketing Science* 38(2):193–225, URL http://dx.doi.org/10.1287/mksc.2018.1135.

Hartmann J, Exner Y, Domdey S (2025) The power of generative marketing: Can generative AI create superhuman visual marketing content? *International Journal of Research in Marketing* 42(1):13–31, ISSN 0167-8116, URL http://dx.doi.org/10.1016/j.ijresmar.2024.09.002.

Heitmann M, Jansen TPJ, Reisenbichler M, Schweidel DA (2025) EXPRESS: Picture Perfect: Engaging Customers with Visual Generative AI. *Journal of Marketing* 0(ja):00222429251356993, URL http://dx.doi.org/10.1177/00222429251356993.

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.

Huang M, Katona Z (2025) Consumer evaluation and visual attribution of realistic AI-generated advertisements. *SSRN Electronic Journal* URL http://dx.doi.org/10.2139/ssrn.5366745.

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 Eco-*

nomics & 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.

Mahdawi A (2025) AI-generated 'slop' is slowly killing the internet, so why is nobody trying to stop it? URL www.theguardian.com/global/commentisfree/2025/jan/08/ai-generated-slop-slowly-killing-internet-nobody-trying-to-stop-it.

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 www.taboola.com/press-release/generativeai.

Thompson SA (2024) Test Yourself: Which Faces Were Made by A.I.? URL 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.

Wedel M, Pieters R (2015) The Buffer Effect: The Role of Color When Advertising Exposures Are Brief and Blurred. *Marketing Science* 34(1):134–143, ISSN 0732-2399, URL http://dx.doi.org/10.1287/mksc.2014.0882.

Zehnle M, Hildebrand C, Valenzuela A (2025) Not all AI is created equal: A meta-analysis revealing drivers of AI resistance across markets, methods, and time. *International Journal of Research in Marketing* ISSN 0167-8116, URL http://dx.doi.org/10.1016/j.ijresmar.2025.02.005.

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.