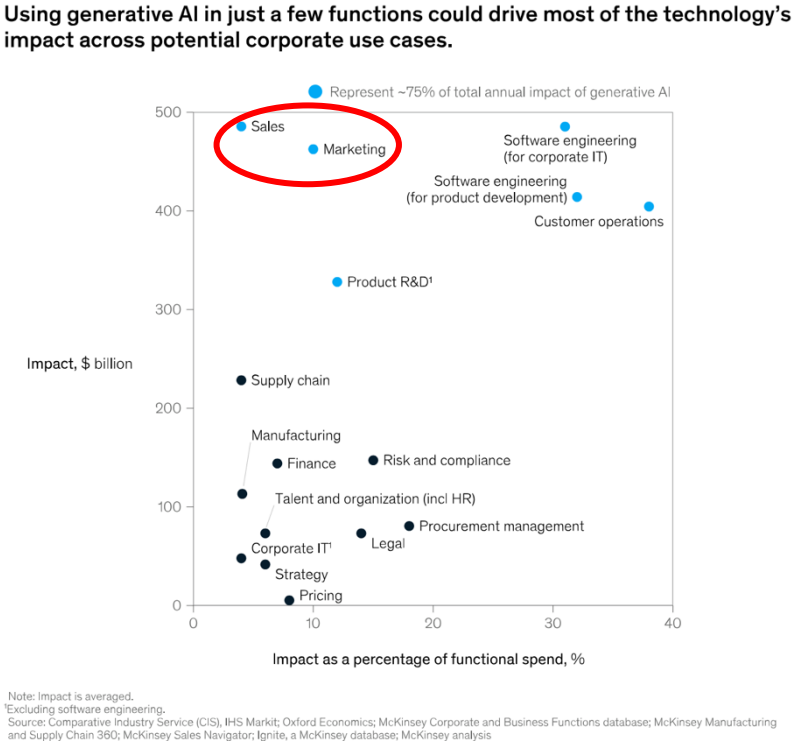
AI in disguise: AI-generated ads outperform human-made ads if they don't look like AI

MSI at the ARF Forum

Taboola · Columbia Business School · Harvard Business School · TUM · Carnegie Mellon Tepper School of Business

Oded Netzer
Columbia Business School

Generative AI disrupts marketing as we know it



Source: McKinsey (2023)

AI is disrupting the advertising industry

Research Questions

- Do advertising creatives generated by AI perform better than human creatives?

- Answering the question using individual A/B tests is problematic
  - Can only provide snapshots into specific application contexts (e.g., Hartmann et al. 2024)

Quiz: Human or AI?

Collaboration with a display advertising platform

We take a bird's eye view and assess a real-life dataset of a leading online ad platform:

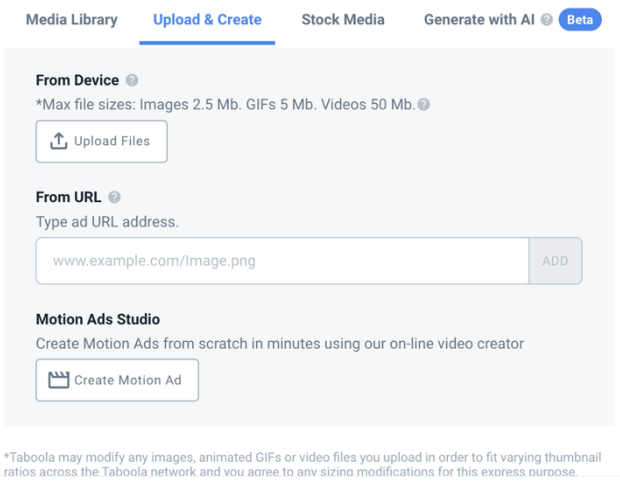
- Platform released generative AI-powered Ad Maker in July 2023
- 2M+ daily ad-level observations across various industries, accumulating 16B+ impressions and 116M+ clicks
- Includes both human-made and AI-generated ads in quasi-experimental setting (29k+ daily ad-level observations)

Taboola

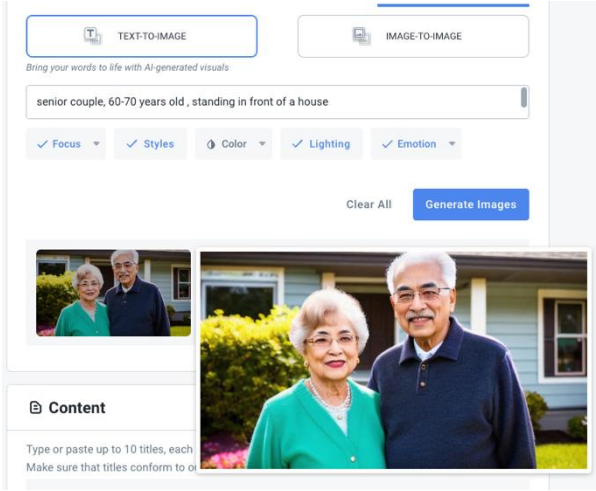
Taboola Columbia Business School Harvard Business School TUM Tepper School of Business

Research context: Gen AI Ad Manager workflow

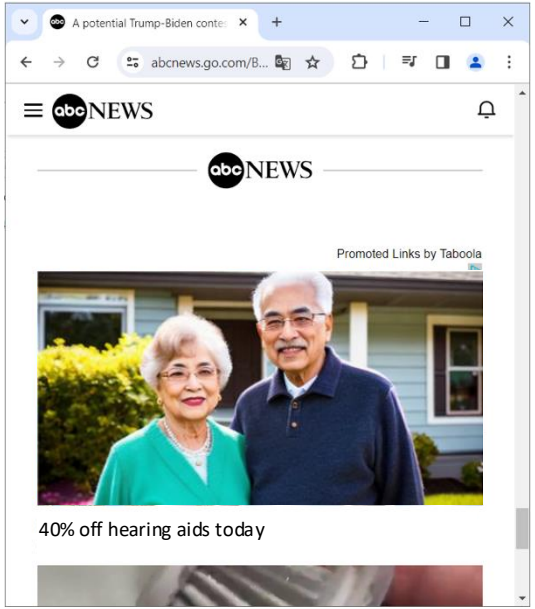
Old world: Human-made ads



New world: AI-generated ads



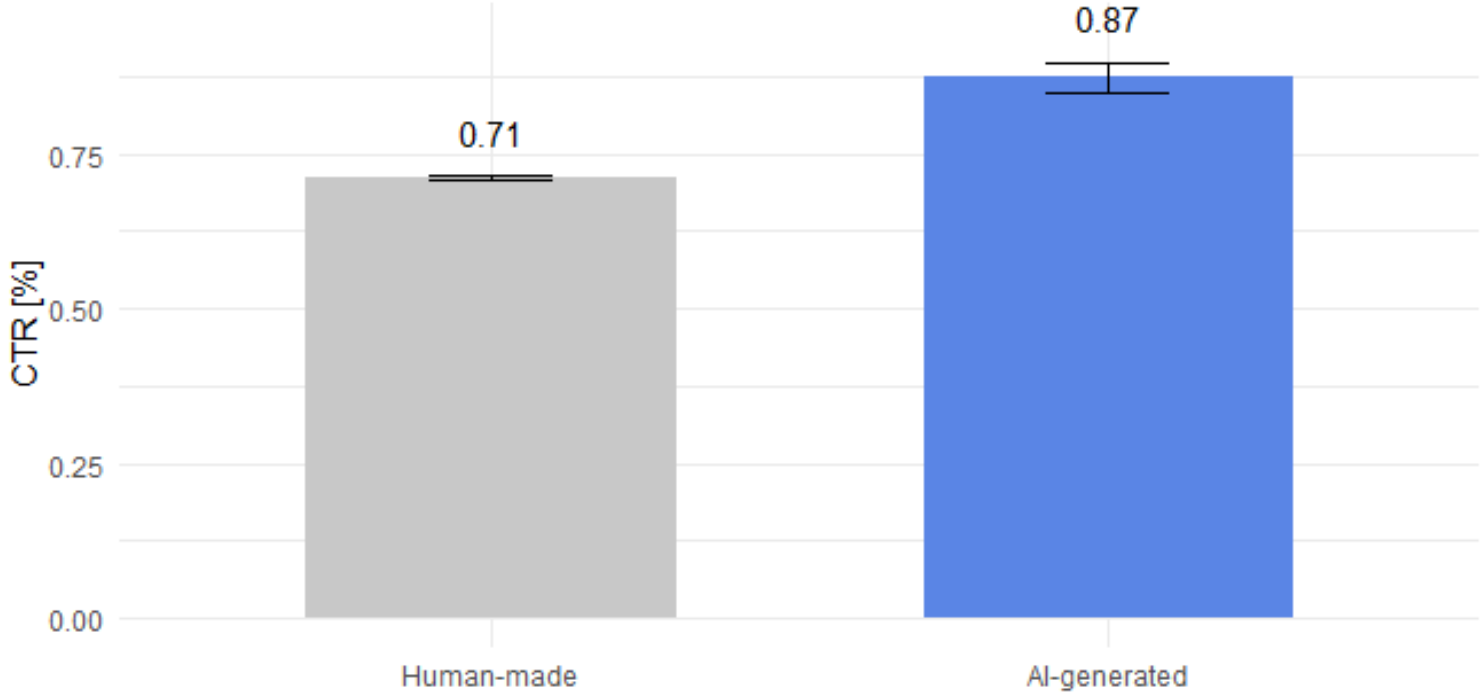
Distribution



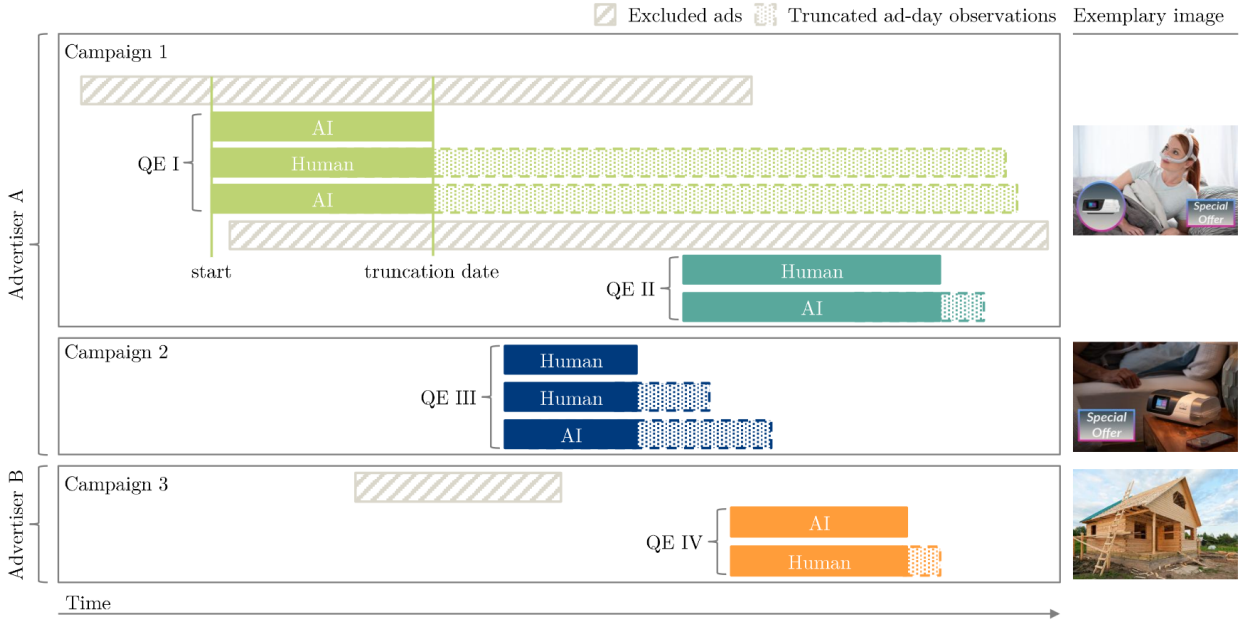
GenAI Ad Maker: A New Way to Create High-Quality Ads Effortlessly

Source: Taboola (2023)

Model-free evidence suggests outperformance of AI-generated images versus human-made ones



Schema of our quasi-experiment design of ads



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.

Each individual "A/B test" results from advertisers' systematic experimentation under controlled conditions

Resulting sample size:

- 29k+ daily ad-level observations
- 4,633 ads
- 202 advertiser

Exemplary ads from same quasi-experiments

Human-made | AI-generated | CTR

Unsold Gazebos Are Selling Almost For Nothing!

Unsold Gazebos Are Selling Almost For Nothing!

New Small Electric Car For Seniors - The Price May Surprise You

New Small Electric Car For Seniors - The Price May Surprise You

Seniors: $200 Bonus On Checking Account is Turning Heads

Seniors: $200 Bonus On Checking A ccount is Turning Heads

Top 5 Best Laptop For Programming in 2023

Top 5 Best Laptop For Programming in 2023

When gradually zooming into our quasi-experiments, AI-generated images seem to perform on par with human-made ads

Previously defined quasi-experiments

Dependent Variable:	CTR				
Model:	All observations (1)	Advertisers with AI & human ads (2)	Campaigns with AI & human ads (3)	AI & human ads start on same day (4)	Right truncation (5)
Variables					
AI-generated image (effect-coded)	.1267*	.1513*	.0494*	.0218	-.0490
	(.0530)	(.0642)	(.0246)	(.0298)	(.0272)
AI-generated caption (effect-coded)	-.1184*	-.1412*	-.1767***	-.0767	-.0748
	(.0576)	(.0688)	(.0470)	(.1122)	(.1085)
Controls caption	Yes	Yes	Yes	Yes	Yes
Controls description	Yes	Yes	Yes	Yes	Yes
Controls image	Yes	Yes	Yes	Yes	Yes
Fixed-effects					
Calendar date	Yes	Yes	Yes	Yes	Yes
Advertiser id	Yes	Yes			
Campaign id			Yes		
Quasi-experiment id				Yes	Yes
Fit statistics					
Observations	2,218,322	190,718	40,044	40,024	29,592
Squared Correlation	.0394	.0751	.0694	.1184	.1965
Pseudo R^2	.9563	.9561	.9799	.9823	.9834
BIC	60,733,901.9	9,398,541.6	834,614.2	744,608.0	515,648.4

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

Human bias against AI and AI-generated content exists

Humans exhibit bias against algorithms and AI-generated content (Dietvorst et al. 2015, Luo et al. 2019) and devalue AI-made art (Horton et al. 2023)

However, **humans perceive AI and human-made content on par in terms of quality and realism** (Hartmann et al. 2025) **and struggle to tell it apart** (Jakesch et al. 2023)

Apart of some high-level consumer reactions (Hartmann et al. 2025), we only have **limited understanding how perceived artificiality of AI-generated content affects real-world consumer behavior**

Sources:
Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114.
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)
Horton Jr, C. B., White, M. W., & Iyengar, S. S. (2023). Bias against AI art can enhance perceptions of human creativity. *Scientific Reports*, 13(1), 19001.
Jakesch, M., Hancock, J.T., & Naaman, M. (2023). Human heuristics for AI-generated language are flawed, *Proc. Natl. Acad. Sci. U.S.A.* 120 (11) e2208839120
Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019) Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science* 38(6):937-947.

Capturing an ad's perceived artificiality ('looks-like-AI')

Is this image human-made OR AI-generated?

Select an option

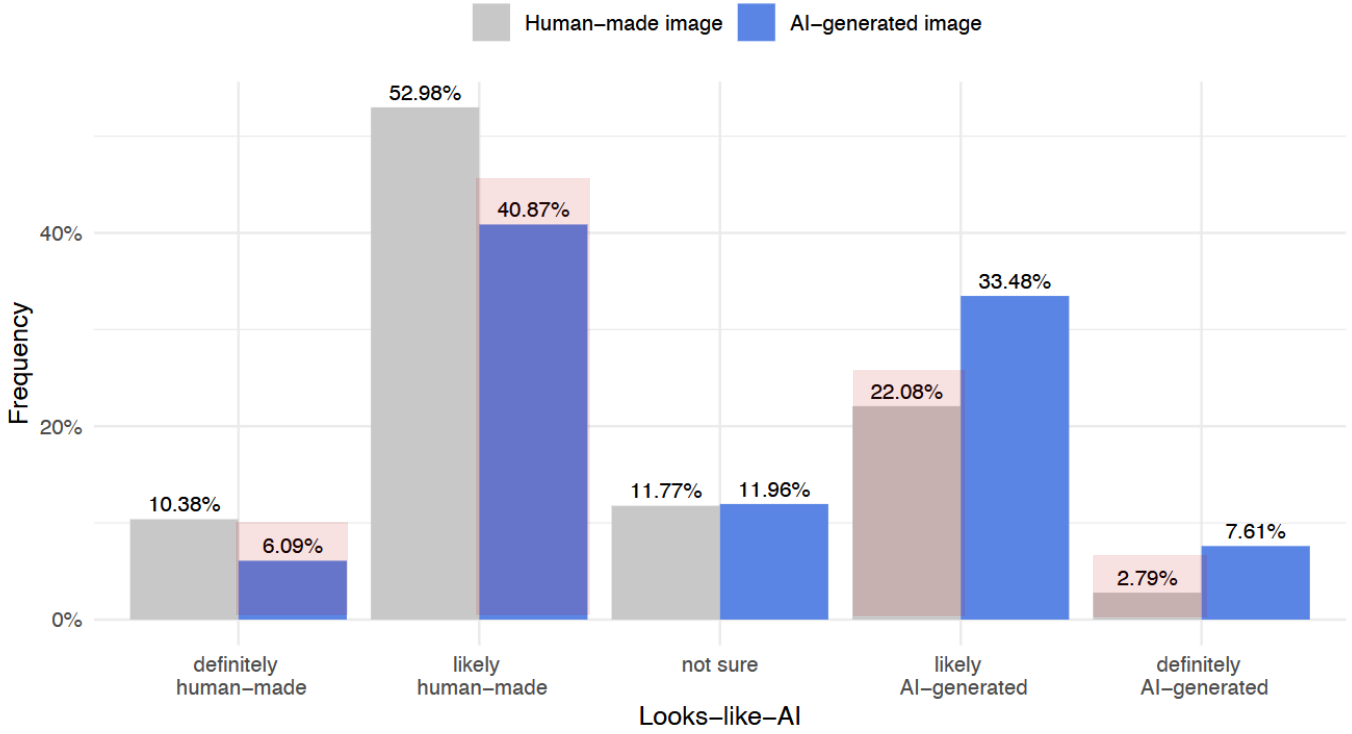
- definitely human-made
- likely human-made
- not sure
- likely AI-generated
- definitely AI-generated

Submit

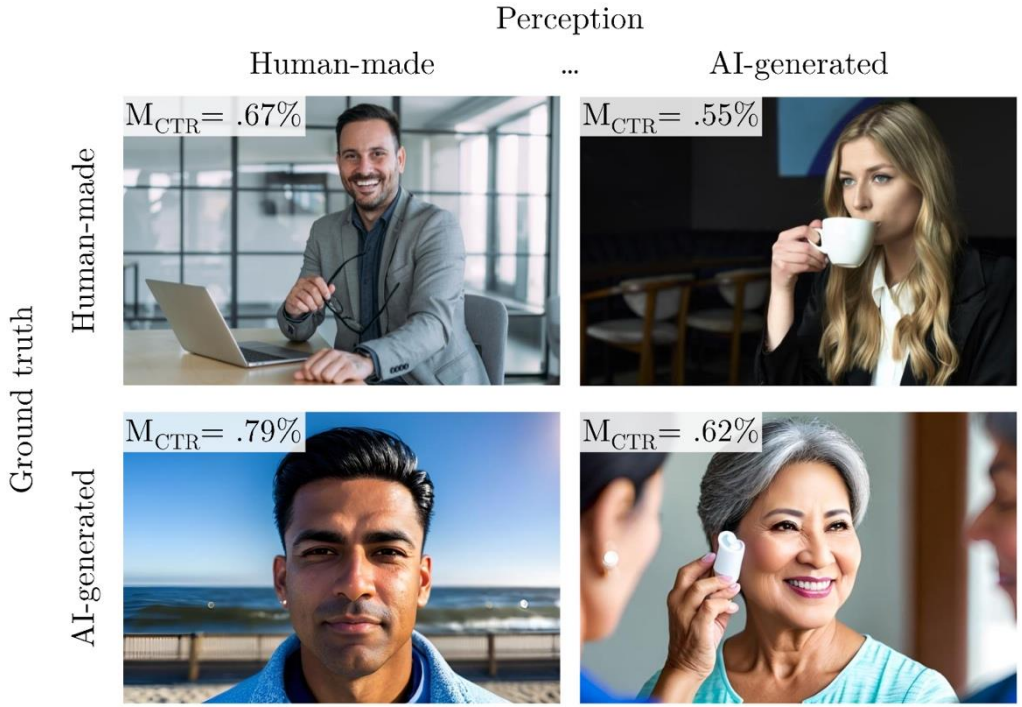
Likert scale adopted from *Jakesch, M., Hancock, J. T., & Naaman, M. (2023). Human heuristics for AI-generated language are flawed. Proceedings of the National Academy of Sciences*

Taboola Columbia Business School Harvard Business School TUM Tepper School of Business

Consumer ratings (blinded) of perceived artificiality of quasi-experimental ads

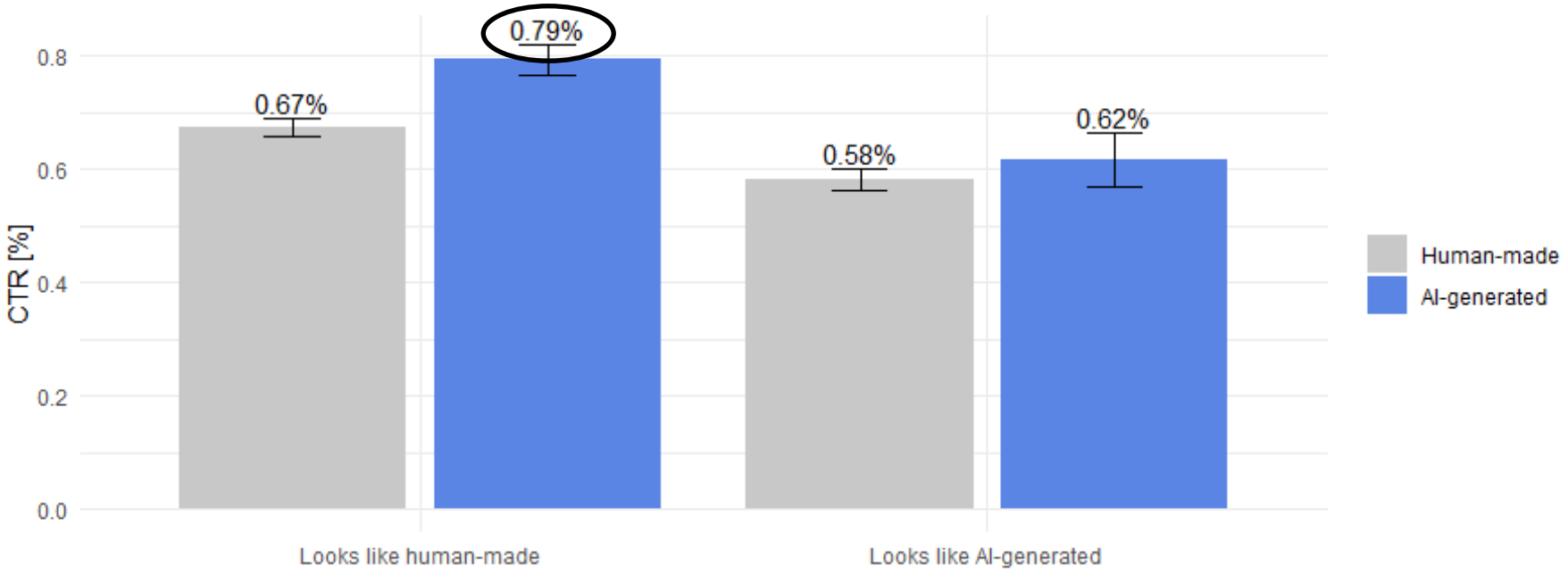


Model-free evidence: AI ads that do not look like AI perform best



Note: CTRs in white boxes indicate model-free ad performance by quadrant

Model-free evidence: AI ads that do not look like AI perform best

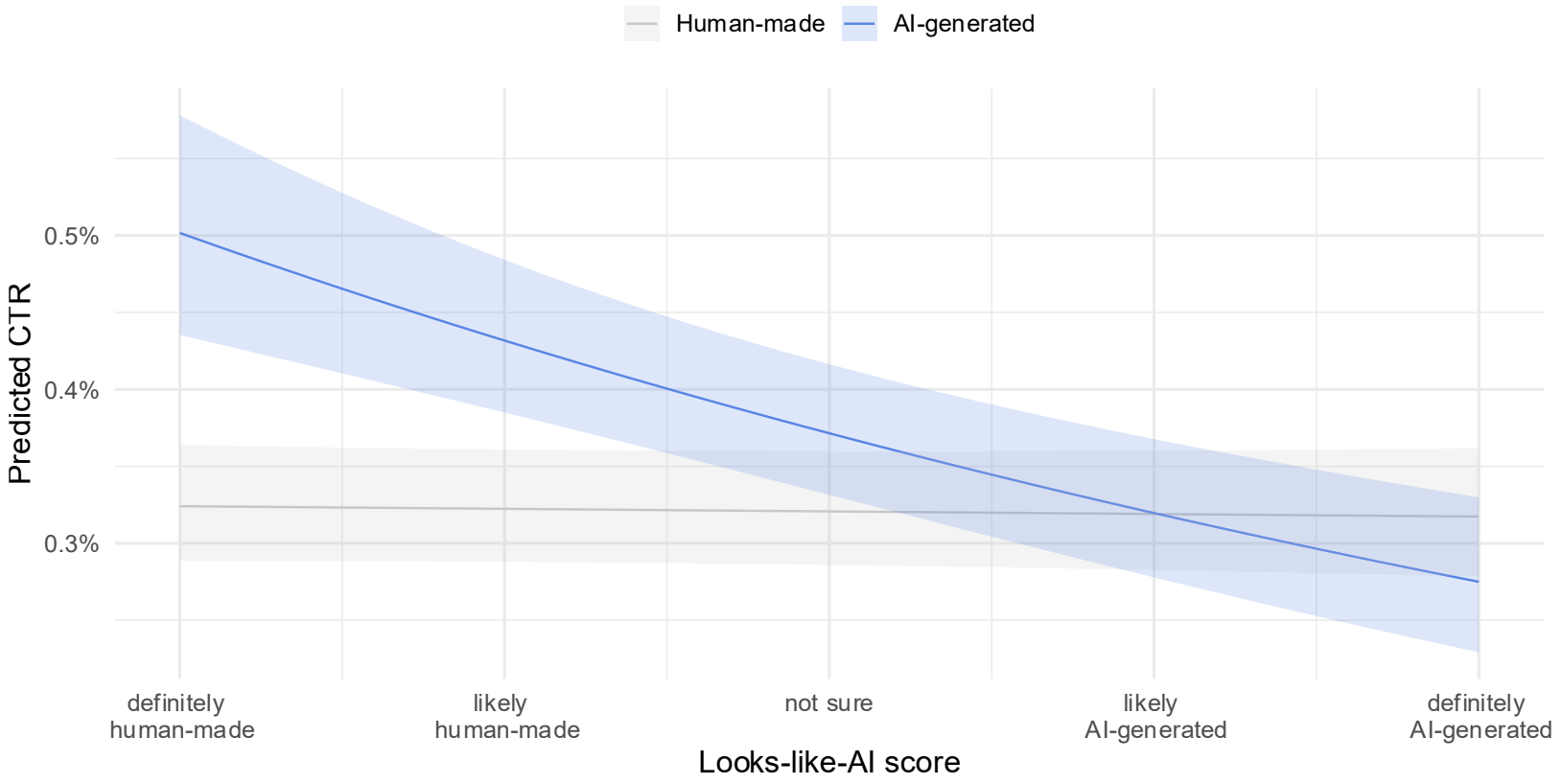


How consumers respond to AI-generated ads (1/2)

Dependent Variable:	CTR				
Model:	All observations (1)	Advertisers with AI & human ads (2)	Campaigns with AI & human ads (3)	AI & human ads start on same day (4)	Right truncation (5)
Variables					
AI-generated image (effect-coded)	.1312**	.1591*	.0684*	.0686+	.0182
	(.0478)	(.0651)	(.0273)	(.0397)	(.0393)
AI-generated caption (effect-coded)	-.1099*	-.1286*	-.1834***	-.0688	-.0633
	(.0551)	(.0645)	(.0444)	(.1155)	(.1073)
looks-like-AI (mean-centered)	-.4760**	-.4467**	-.3241*	-.2941	-.4880*
	(.1646)	(.1651)	(.1391)	(.2240)	(.2405)
AI-generated image (effect-coded) × looks-like-AI (mean-centered)	-.4383**	-.5772**	-.3103*	-.4676**	-.4744**
	(.1434)	(.1828)	(.1217)	(.1774)	(.1786)
Controls caption	Yes	Yes	Yes	Yes	Yes
Controls description	Yes	Yes	Yes	Yes	Yes
Controls image	Yes	Yes	Yes	Yes	Yes
Fixed-effects					
Calendar date	Yes	Yes	Yes	Yes	Yes
Advertiser id	Yes	Yes			
Campaign id			Yes		
Quasi-experiment id				Yes	Yes
Fit statistics					
Observations	2,218,322	190,718	40,044	40,024	29,592
Squared Correlation	.0394	.0752	.0693	.1188	.1976
Pseudo R^2	.9563	.9562	.9780	.9823	.9835
BIC	60,716,442.3	9,380,317.8	831,138.2	742,102.4	512,750.9

*Signif. Codes: ***: .001, **: .01, *: .05, +: .1*

How consumers respond to AI-generated ads (2/2)



Assessing perceptual, structural and content-related visual features

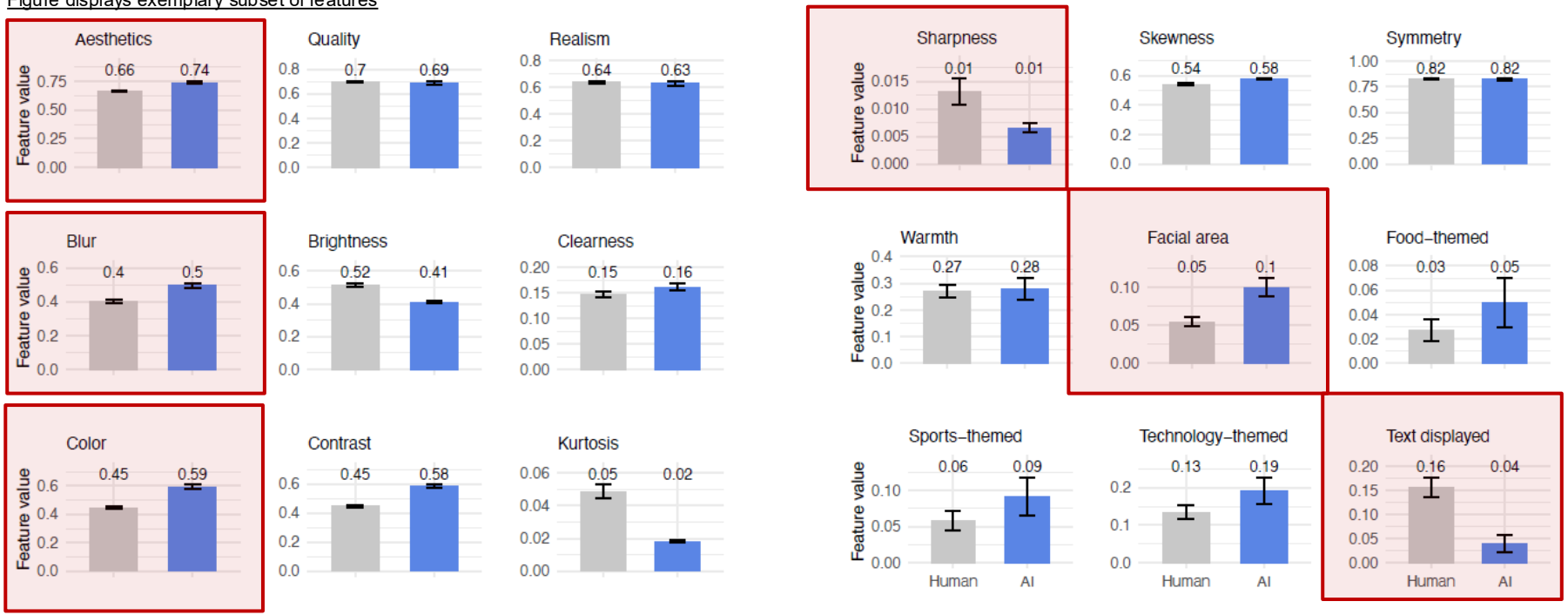
Feature	Description	Source
Perceptual		
Aesthetics	Perceived image aesthetics	1
Quality	Perceived image quality	4
Realism	Perceived image realism	4
Structural		
Blur	Perceived blur annoyance	2
Brightness	Mean grayscale pixel intensity	3
Clearness	Average gradient magnitude of geometric mean from horizontal and vertical Sobel operator gradients	1
Color	Colorfulness := mean(S_i) + std(S_i), where S_i is the S values in HSV color space for image i	3
Contrast	Standard deviation across all pixels in grayscale	3
Kurtosis	Kurtosis of pixel intensities distribution	3
Sharpness	Variance of Laplacian operator (i.e., degree of change in image edges and details)	3
Symmetry	Normalized norm of difference between left and right image halves	3
Warmth	Proportion of warm versus cold hues in HSV space	3
Content		
Facial area	Relative image area covered by faces' bounding boxes	4
Food-themed	Aggregates food-related COCO labels in a dummy variable	5
Sports-themed	Aggregates sport-related COCO labels in a dummy variable	5
Technology-themed	Aggregates tech-related COCO labels in a dummy variable	5
Text displayed	Indicates if text is displayed	5

Sources: 1 - Talebi and Milanfar (2018), 2 - Crété-Roffet et al. (2007), 3 - Zhang et al. (2022), 4 - Hartmann et al. (2024), Beichert et al. (2024), 5 - Lin et al. (2014), Web Appendix Figure 2 displays exemplary images for low and high feature values

+ 6 additional structural features:
- Aspect ratio
- Depth of field (hue)
- Diagonal dominance
- Rule of thirds
- Shannon entropy
- Visual complexity

Which visual feature are prominent in AI images?

Figure displays exemplary subset of features



Notes. For ease of comparability within the figure, the feature values in this plot are normalized ($x \in [0, 1]$)

Human-made image AI-generated image

Relationship of looks like AI and image features in AI imagery

Dependent Var.:	Perceived artificia.. pred_looks_like_ai_01	AI-generated image AI-generated image
AI-generated image	0.0465*** (0.0121)	
Aesthetics	0.0243*** (0.0047)	1.342*** (0.2038)
Quality	-0.0006 (0.0041)	0.1379 (0.1753)
Realism	-0.0057 (0.0040)	0.0896 (0.1591)
Blur	0.0525 (0.0373)	-5.632 (4.792)
Brightness	0.0059 (0.0052)	-0.7666** (0.2950)
Clearness	-0.0187* (0.0081)	0.9678* (0.3758)
Color	0.0257*** (0.0051)	0.8136** (0.2721)
Contrast	-0.0017 (0.0051)	1.902*** (0.2587)
Kurtosis	0.0076 (0.0056)	1.098 (0.8253)
Sharpness	-0.0321 (0.0370)	3.202 (4.754)
Symmetry	0.0061 (0.0047)	0.2452 (0.2699)
Warmth	-0.0121** (0.0042)	-0.4295** (0.1614)
Aspect ratio	-0.0038 (0.0043)	-0.3695+ (0.2099)
Depth of field	0.0117** (0.0041)	1.088*** (0.1785)
Diagonal dominance	0.0034 (0.0040)	-0.4675* (0.1819)
Rule of thirds	-0.0051 (0.0040)	0.0810 (0.1707)
Shannon entropy	-0.0122+ (0.0068)	5.626*** (1.082)
Visual complexity	0.0074+ (0.0044)	0.4662* (0.2375)
Facial area	-0.0169*** (0.0043)	0.8117*** (0.1690)
Food-themed	0.0231 (0.0186)	1.158 (1.104)
Sports-themed	-0.0052 (0.0147)	0.8930 (0.6016)
Technology-themed	-0.0309** (0.0116)	0.0330 (0.4416)
Text displayed	0.0068 (0.0132)	-0.9931+ (0.5485)
Fixed-Effects:	----------	----------
Advertiser id	Yes	Yes
Family	OLS	Logit
S.E. type	Heteroskedastic.-rob.	Heteroskedas.-rob.
Observations	1,751	1,697
Squared Cor.	0.27140	0.77223
Pseudo R2	-0.34881	0.73415
BIC	-247.04	2,134.0

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1

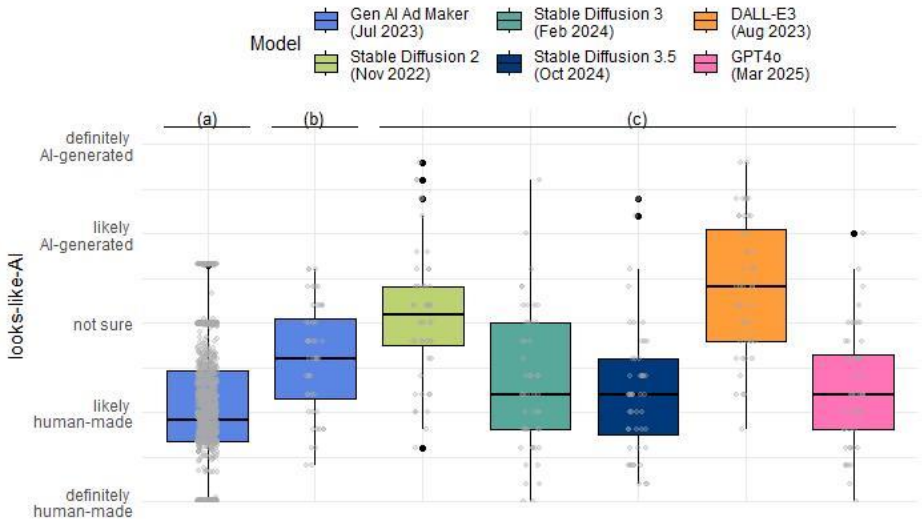
β>0 β<0

Humans identify AI:
More colorful images look like AI and AI generates more colorful images

Humans **can't** identify AI:
Small faces look like AI **but** AI actually generates larger faces

Note: Continuous variables are z-standardized

Newer AI models still exhibit looks-like-AI heterogeneity



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

Conclusion

Summary

Real-world dataset leveraging a quasi-experimental design showed, that **AI images**:

- **typically reach human-level performance**
- **outperform human-made images if they do not look like AI**
- **cover more image area with faces** – which makes them look less like AI

Implications

Use AI-generated imagery to **increase online banner ad effectiveness**

Still: Ensure careful consideration of structural and content-related (e.g., leverage images with more facial area)

Source: DALL-E3

It does not look like AI – but is it AI in disguise?

Thank you!

Oded Netzer

Columbia Business School

onetzer@gsb.columbia.edu

It does not look like AI – but is it AI in disguise?

Thank you!

Oded Netzer

Columbia Business School

onetzer@gsb.columbia.edu

Thank you!

BACKUP

Quality and Realism were labelled via MTurk

Please answer the questions based on the image displayed

Question 1/2:

The visual elements of the ad are realistic

Definition of "realistic": accurately representing what is natural or real

1	2	3	4	5	6	7
○	○	○	○	○	○	○
strongly disagree	disagree	somewhat disagree	neither agree nor disagree	somewhat agree	agree	strongly agree

Question 2/2:

Give a score to an image on a scale of 1-7 on its aesthetic quality where 1 is "very bad" and 7 is "excellent"

Detailed instructions (click to collapse)

We are interested in assessing the aesthetic quality of images. Give an aesthetic score to an image on a scale of 1-7 on how beautiful/pleasing/visually appealing you think the image is, where 1 is "very bad" and 7 is "excellent". Here is a short non-exhaustive guide to judging images.

Reasons for a rating closer to 7:

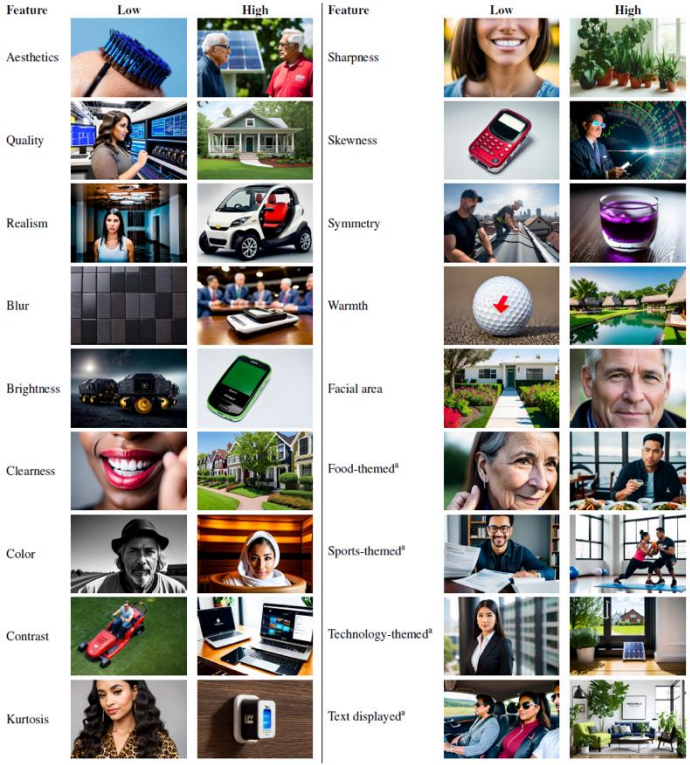
- looks good
- visually pleasing
- well-lit
- clearly shows features
- conveys well what the scene may look like in person

Note: The score should be given on aesthetic quality of an image, regardless of its topic

1	2	3	4	5	6	7
○	○	○	○	○	○	○
very bad	bad	below average	average	good	very good	excellent

Source: Hartmann et al. (2024)

Exemplary AI-generated images for each feature



Notes. Left image always shows low feature value while right column shows high feature value (e.g., brightness for left image is 30.37 and for right image is 223.53 in mean grayscale pixel intensity). [a] indicates dummy variables, where low = 0 and high = 1.

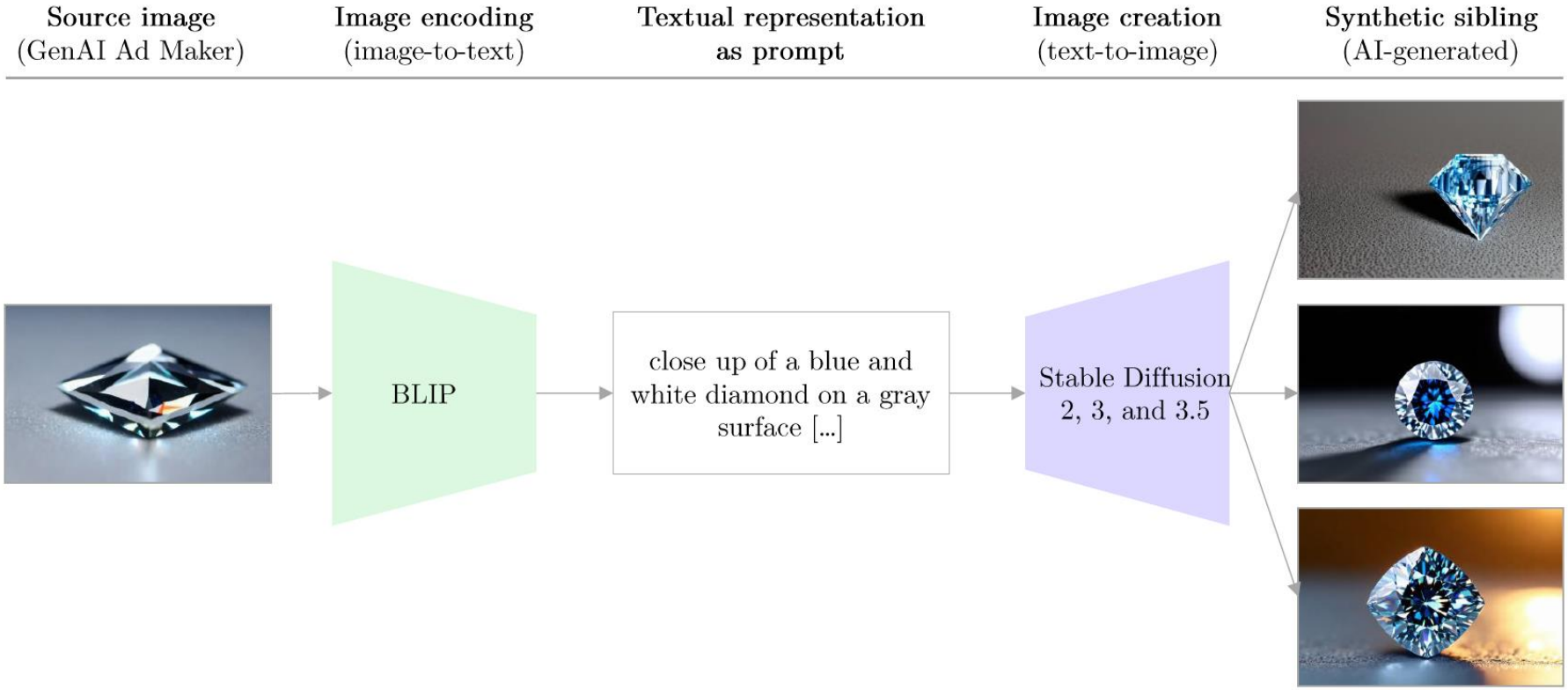
Robustness check: Self-selection of the human-made ads

Dependent Variable:	CTR
Model:	(1)
Variables	
Ad in quasi-experiment	.1069
	(.1157)
AI-generated caption	-.1329
	(.1697)
Controls	
CPM	Yes
Verbal features of caption	Yes
Verbal features of description	Yes
Visual features of creative	Yes
Fixed effects	
Advertiser	Yes
Calendar date	Yes
Fit statistics	
Observations	117,952
Squared correlation	.33407
Pseudo R^2	.9673
BIC	5,925,481

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

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

Image regeneration approach with newer models



Source. Hartmann et al. (2024)