The Introduction of Human and AI-Powered Virtual Influencers on Social Media Marketing

by Serim Hwang

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Marketing

at the Tepper School of Business

Carnegie Mellon University

2023

Dissertation Committee:

Kannan Srinivasan (Chair)

Xiao Liu

Alan Montgomery

Hui Li

Yan Huang

Acknowledgements

First and foremost, I was privileged to work with my advisor, Professor Kannan Srinivasan. He sparked my thought process, fueled me how to find great research questions, and we, at last, illuminated the mounted treasures of digital and influencer marketing. He has relentlessly sharpened his insights on research and I was marveled by how the past few years have been evolving as the projection of his insights. Even though everyone in our field knows so well, to name a few of his enlightening identities, my advisor Kannan is a great leader, mentor, and a game changer. Naturally, his groundbreaking identities made our CMU Marketing and BT solid and stronger. His contribution will everlast as I and his great students will keep pursuing leading and innovative research.

I will forever be indebted to my advisor, Professor Xiao Liu. Xiao is incredibly smart, hard working, and thus super productive, beyond a doubt! At the same time, she is well-mannered, elegant, and gently graceful to her collaborators. I have always admired and adored her, her mesmerizing ideas, her state-of-the-art research innovations, and her professional work ethics. In fact, it was truly one of my greatest honors to work with this world-class outstanding researcher. I firmly believe that she will be the next leader in our field. Xiao+Serim collaboration was full of fun, and Xiao genuinely inspired me to grow higher, learn faster and stay brighter. I will always try my best to support Xiao and to become one of her many successful students.

I want to thank two professors who shed light on my PhD journey, Professor Alan Montgomery and Professor Hui Li. Alan thankfully gave me many learning opportunities and trained me hard on how to think carefully. Under his mentorship, I have developed a keen understanding of financial decision-making structures and learned to navigate potential endogeneity concerns. I will try my best to carry forward the rigorous thinking skills that I have been trained in by Alan.

I feel so thankful for Professor Hui Li's great research support. Her mentorship has been invaluable in sharpening my research skills and expanding my horizons. She thoughtfully and patiently taught me how to build structural models, generously provided great resources, and helped me to think critically about the theoretical frameworks for empirical problems. I was so fortunate to receive her high-quality trainings. I believe these two professors will make our marketing research shining!

Special thanks to Professor Yan Huang for accepting my dissertation committee invitation and being one of my precious faculty committee team members. She was kind enough to make my invitation decision so easy through her excellent research achievements. I will always be longing for her next great research.

For a special note, I am so grateful for Professor Param Singh, Professor Nitin Mehta, and Professor Vineet Kumar who kindly helped me while I was passing my job market process.

I want to deliver my warmest thank-you message to the following marketing professors, Professor Fred Feinberg, Professor Anocha Aribarg & Professor Kee Yeun Lee. They are truly my lifesavers. Without their big-hearted support on applying to the marketing PhD programs, I would not have discovered this exciting new career.

It was my true honor to learn and get positive influence from our CMU marketing professors: Peter Boatwright, Tim Derdenger, Jeffrey Galak, Christopher Y. Olivola, Tong (Joy) Lu and Peter Stüttgen. It was truly amazing and grateful to cross paths with the following professors at Tepper:

Soohaeng Cho, Karam Kang, Taewan Kim, Sunkee Lee, Dokyun Lee, R. Ravi, Stephen E. Spear, David Childers, Serra Boranbay Akan, and Ross Doppelt.

I want to deliver my dearest thank-you messages to Lawrence & Laila. You will never know how encouraging and empowering it has been to have both of you during the last six years.

My academic brothers and sisters: Nikhil Malik, Shunyuan Zhang, Zijun (June) Shi, Yijin Kim, and Francisco Cisternas. Nikhil has been my great mentor who continuously gave me wonderful advice. Shunyuan has always been my great sister. She has supported me to do more impactful and interesting research. June kindly offered me great guidance in learning various research resources. Yijin thoughtfully led my way to smoothly start my PhD journey. Francisco thankfully provided me research support and advice. Outside of the Tepper, yet still in CMU, we have three great colleagues: Emaad Manzoor, Runshan Fu, and Siddhartha Sharma. We all used to have such a fun time together at conferences. In our next conference travels, I will take the role of a gourmet food tour guide as what Emaad used to do:) In retrospect, I have received way too much support from my academic sisters and brothers. I hope this supportive CMU culture to be long run!

I want to deliver the sincere thanks to my cohorts: Sagnik Das, Xinying (Violet) Chen, Thomas Lavastida, Yuyan Wang, Jingyi (Kyra) Gan, Zahra Ebrahimi, Musa Eren Celdir, Ozgun Elci, Melda Korkut, Majid Mahzoon, Diana Mikhail, Su Jia, Benjamin Ostrowski, and Timothy Hyde. Growing up in this interdisciplinary academic culture has been full of pleasant surprises. Learning about their research and sharing our thoughts together were of great help to improve my research exponentially. You all gave me unforgettable memories both intellectually and as a human. If I am lucky enough, I would love to meet all of you once again for the rest of my life. And I want to thank my dearest Korean cohorts: Jinwoo Kim, Jisoo Park and Sae-Seul Park. I am so thankful to receive their warm and kind help and support during my PhD journey.

I feel so thankful to the following senior Ph.D. graduates for giving me great advice: Santiago Tellez, Franco Berbeglia, Neda Mirzaeian, Bo Yang, Joachim Talloen, Ziye Tang, Xuege Zhang, Sang Wu, Zhaoru (Aluna) Wang, and Ali Polat. Thank you so much for Korean PhD students as well: Hyun Hwang, Eungsik Kim, and Ki-Won Haan.

I am also thankful to our brilliant fellow marketing Ph.D. students who kindly and gently made me feel supported: Samuel Levy, Behnam Mohammadi, Xiaohang (Flora) Feng, Ziyan (Angela) Xiao, Yuan Yuan, Jieqiong (Julie) Wang, and Jaymo Kim. I will wait for the moment that I will be surprised by how these CMU Tepper marketing scholars will give positive big impacts and enrich our research community!

I was fortunate to interact with the following great fellow students: Jaepil Lee, Liying Qiu, Qiaochu Wang, Yin (Lavender) Yang, Yanhan (Savannah) Tang, Xiaonan Hong, Zhaoqi Cheng, Neha Neha, Minhyuk Nam, Jae Yeon (Jane) Pyo, Jenny Oh, Yong Joo Won, Genaro Basulto, Arnav Sood, Allen S. Brown, Martin Michelini, Jason Gates, Anthony Karahalios, Daniel de Roux, Sunan Qian, Preetika Basak, Rudy Zhou, and Mik Zlatin.

Thanks to my job market cohorts and friends from Michigan: Sherry He and Yunfei (Jesse) Yao. It was my great pleasure to interact with these job market cohorts as well: Ian Weaver and Xinyao Kong. Also, thank you Fanglin Chen, Lu Zhang, Gaomin Wu, and Chenshuo Sun, for giving me great research and lifelong memories in Manhattan! Working with NYU Marketing/IS group was such a cheerful and fun experience that I would never forget! Thank you very much for our much-esteemed Xiao for inviting me to Manhattan and leading us to awesome achievements!

I feel so thankful to our SKK GSB professors who have greatly supported me to successfully complete my dissertation for the last four months by making my office earlier and making me feel like at home: Professor & Dean Eric Chuan Fong Shih, Professor Kyeongheui Kim, Professor Minjung Koo, Professor Nah Lee, Professor Xing Zhang, Professor John Van Den Berghe, Professor Andrew Chunwon Yi, Professor Raunaq Pungaliya, Professor Suil Pae, Professor Aleksios Gotsopoulos, Professor Youngju Nielsen, Professor Frank Huettner, Professor Adam R. Castor, Professor Kim Claes, and Professor Chengwei Wang.

I would like to express my heartfelt gratitude to my amazing PhD sisters: Jimi Kim, Yubin Kim, Jungyoon Jang, and my dearest cousin Hyejin Sung. I would also like to extend my sincere thanks to my CMU Korean CS/Stat/Engineering/Architecture PhD brothers: Byungsoo Jeon, Byungjoo Ahn, Jinmo Rhee, Heejong Bong, Sangsuk Lee, Namyong Park, Kijung Shin, and Hoon Oh. They have taken care of me as a little sister and I feel grateful to my sisters and brothers for their friendship, support, and the positive impact they have had on my life. They genuinely influenced me to happily follow their career and life paths.

Finally, I am blessed to have my top priorities and serotonin boosters - my dad Youngdo Hwang, my mom Jihyun Kang and my little sister Sehyun Hwang.

Dad, you are like a calm, deep blue ocean where nothing happens no matter how many stones I throw. Your belief in me gives me beautiful strength. Everytime I see you cry at the airport, I deeply feel how much you love me and feel sorry dad. I already came to realize it is impossible for me to work as powerfully and productively as you do. Nevertheless, I will always try to think in mind of how you would have done better in my shoe. Thank you for being my dad! I love you so much, Abba!

Thank you very much for your warmest support, my wonderful and sweetest mom! Your trust, care, and prayers always cheer me up and empower me to work harder and become a good person. Listening your stories is always refreshing, and it reinforces me in various directions! Your wisdom and life lessons have guided me to the right way that I can cherish and be thrilled with determination. Thank you for being my mom and love you so much, Umma!

My so cute lovely little sister, Sehyun. You have been my adorable inspiration and a great joy! While it has been many years that I could not see all your dazzling moments, from now on, I will do my best to support your dreams, Sennie! Unni will always be on your side. Do whatever you want! Love you, my bestie Sehyun!

Lastly, over the past six years of study, it was so delightful to be a student at this great university! Thank you dear, Carnegie Mellon University!

Table of Contents

1. Voice Analytics of Online Influencers

ABSTRACT

In the rapidly expanding influencer marketing sector, positive consumer sentiment towards sponsored content plays a crucial role in the success of sponsoring brands and influencers. However, research shows that consumers are more likely to react negatively to sponsored videos than non-sponsored ones. We investigate how influencers can potentially diminish the negative impact of sponsorship on consumer sentiment by modulating their vocal characteristics in sponsored videos compared to non-sponsored videos. We use three causal identification strategies: instrumental variables, fuzzy regression discontinuity, and two-way fixed effects. To extract multi-modal 3V features (voice, visual, and verbal), we employ deep learning methods. We conclude that influencers significantly reduce the average loudness of their voices in sponsored videos relative to non-sponsored ones, and this difference in average loudness partly alleviates the negative impact of sponsorship on consumer sentiment. This vocal strategy of lowering loudness is more likely to be employed by influencers with smaller followings, as opposed to those with larger followings. We discuss the implications for various stakeholders, including influencers, brands, consumers, and regulatory bodies.

Keywords: influencer marketing, social media advertisements, voice analytics, affective computing, image recognition.

INTRODUCTION

"It is not what you say, but how you say it." – Mae West

The influencer marketing industry has witnessed significant growth, and it is expected to reach $15 billion by 2022, with a growth rate of 42% in 2021 (Globe Newswire 2021). Brands such as Sigma Beauty have successfully leveraged influencer marketing, generating $25 million in sales through sponsoring beauty influencers on YouTube (Octoly 2015). However, concerns about transparency have led to the Federal Trade Commission (FTC) mandating that influencers disclose their relationship with brands when endorsing products on social media (FTC 2019).

The Federal Trade Commission's disclosure requirement poses a significant risk for influencers as it may result in consumers questioning the authenticity of sponsored videos and responding negatively to the video, influencer, or sponsored brand. Despite the importance of authenticity, few studies have examined how influencers can convey it in sponsored videos and increase sales of the sponsored product. We investigate how influencers may use their vocal characteristics, such as loudness and pitch, which are salient features in almost every video that they can control, to enhance authenticity and generate positive consumer sentiment. While brands often dictate the script to control the influencer's verbal narrative, we highlight the role of vocal modulation strategies, which can be independently employed by influencers to improve consumer response.

We address three research questions related to vocal modulation in sponsored videos by influencers. First, how do influencers modulate their vocal characteristics in sponsored videos compared to non-sponsored ones? Second, what is the impact of vocal characteristics on consumer sentiment, and how does the difference in vocal characteristics between sponsored and non-sponsored videos affect this impact? Third, which characteristics of influencers and videos predict the use of vocal modulation strategy in sponsored videos?

In this research, we investigate the impact of influencer marketing in the global cosmetic products industry, which has been valued at $532 billion as of 2019 (Biron, 2019). To extract vocal, visual, and verbal characteristics from videos, we employ deep learning techniques. We argue that voice is the most manipulable characteristic for beauty influencers, as it would be atypical to alter body language while applying makeup (Krause, 2020), and brands generally provide a strict verbal narrative (Gents Scents, 2018). Consequently, we control for visual and verbal characteristics, which are deemed crucial aspects of communication in previous research (Mehrabian, 2017; Batra, 2019), while exploring the extent to which influencers' vocal modulation strategies moderate the impact of sponsorship on consumer sentiment. Specifically, we examine five vocal characteristics that have been demonstrated to influence voice perception in the Automated Speech Recognition (ASR) and affective computing literature (Scherer et al., 1973; Eyben et al., 2010; Schuller et al., 2019): mean intensity (average loudness), mean F0 (average pitch), the zero-crossing rate (loudness variability), standard deviation of F0 (pitch variability), and voicing probability (talking duration). Furthermore, we use deep learning-based computer vision methods, including a Generative Adversarial Network (GAN; Li et al., 2018), to account for visual characteristics.

The identification of the effects of sponsorship, vocal characteristics, and their interaction on sentiment in influencer marketing is complicated by endogeneity concerns. Positive recognition of an influencer in online media could result in a "popularity shock," increasing their chance of receiving sponsorship offers and receiving more positive sentiment on new videos. To address these concerns, we employ three identification methods. First, we use the IV-PSM approach, employing propensity score matching to identify non-sponsored videos that are similar to sponsored videos in the sponsorship history and influencer/video characteristics. Next, we define two instrumental variables (IV) - the number of sponsorships from the same brand and the parent company of the brand in the same week - that should positively predict treatment (sponsorship) without affecting the outcome (consumer sentiment). Third, we use the fuzzy regression discontinuity (FRD) approach to compare the influencer's first sponsored video for a brand with the non-sponsored video immediately preceding it. Finally, we use the two-way fixed effects (TWFE) approach to adjust for unobserved influencer-specific and time-specific confounders simultaneously. These approaches aim to overcome endogeneity concerns and provide accurate estimates of the effects of sponsorship and vocal characteristics on consumer sentiment.

We find that influencers significantly reduce their average loudness in sponsored videos by 6.9% compared to equivalent non-sponsored videos. This decrease in average loudness mitigates the negative effect of sponsorship by nearly 49%. Specifically, a one standard deviation decrease in the loudness of sponsored videos leads to an increase in consumer sentiment by 0.0026 (-0.0027 + 0.0053 = 0.0026), which dilutes the decrease of 0.0059 in consumer sentiment by approximately 49% in the context of sponsored videos. This is possibly due to the perceived credibility of moderate loudness compared to high loudness (Chebat et al., 2007). Thus, decreasing average loudness represents an effective strategy for influencers to mitigate the potential harm of sponsorship. Further analysis of influencer-specific characteristics reveals that influencers having fewer number of followers are associated with decreased loudness in sponsored videos.

The present research addresses a gap in the literature by examining the influence of vocal characteristics in influencer marketing, an area that has received limited scholarly attention. Along with our findings on this topic, this paper also provides a novel approach for extracting and analyzing marketing-relevant information from diverse sources of unstructured data. As such, we believe our research contributes to the emergent body of literature on the influencer marketing industry, and we hope that it serves as a timely and valuable contribution to this field.

LITERATURE REVIEW

We make a novel contribution to the existing body of literature by investigating the impact of vocal characteristics on seller persuasion within the context of influencer marketing, thereby bridging the gap in the literature on audio analytics and marketing research.

Vocal Characteristics and Persuasion

We aim to contribute to the existing literature on vocal characteristics and seller persuasion by exploring a comprehensive set of vocal features and identifying the characteristics that can effectively mitigate the negative effects of sponsorship on consumer sentiment. Although vocal characteristics are widely employed for persuasion in various businesses, their efficacy can differ depending on the business context. For instance, prior research suggests that higher loudness is linked to lower refusal rates in telephone surveys (Oksenberg et al. 1986), while moderate loudness is perceived as more credible than high loudness in banking telemarketing (Chebat et al. 2007). Additionally, high loudness variability is seen as more attractive (Zuckerman and Driver 1989). A shorter talking duration has been shown to correlate with higher sales performance in direct selling (Peterson et al. 1995). Higher pitch is linked to attractiveness when selecting mates (Fraccaro et al. 2011); however, women with lower-pitched voices are perceived as more trustworthy in leadership roles (Klofstad et al. 2012). Furthermore, election winners tend to have lower-pitched voices than opposing candidates (Banai et al. 2017), while higher pitch variability is associated with lower refusal rates in telephone surveys (Oksenberg et al. 1986). These findings suggest that vocal characteristics can be a valuable tool for persuasion, but their effectiveness is contingent on the business context and specific vocal features employed.

We investigate the impact of five primary vocal characteristics (average loudness, average pitch, loudness variability, pitch variability, and talking duration) on consumer sentiment in the context of social media influencer marketing. Prior literature has primarily examined the effects of vocal characteristics on consumer behavior in controlled experimental settings (Zuckerman and Driver 1989; Peterson et al. 1995; Chebat et al. 2007; Fraccaro et al. 2011), and has overlooked the application of these findings in real-world marketing scenarios such as sponsored influencer videos. One notable exception is the literature by Wiener and Chartrand (2014), which examined the impact of voice quality on consumer purchasing intentions. We extend the literature by leveraging the latest machine learning methods, including affective computing techniques such as voice extraction and sentiment detection. Our findings have important implications for the management of social media influencer marketing campaigns, and contribute to the frontier of marketing research.

Audio Analytics in Marketing

Our research makes an important contribution to the audio analytics in marketing literature by focusing on voice as a specific type of audio sound. While existing marketing literature has explored the emotional impact of music in advertising (Fong et al., 2021) and examined the effect of energy level on ad tuning (Yang et al., 2022), we advance this research by examining how changes in human vocal characteristics, specifically in the context of creating sponsored videos, affect consumer sentiment. By filling this gap in the literature, our paper extends the current knowledge of audio analytics and its implications for marketing strategy.

Influencer Marketing

Our research adds to the growing body of literature on influencer marketing by offering a novel perspective on influencers' vocal strategies. While prior studies have established that cues of message manipulation can negatively affect brand evaluations in advertising (Kirmani and Zhu 2007), our focus is on the crucial role of trust and credibility in the influencer marketing context. Specifically, we investigate the extent to which an influencer's vocal characteristics impact their perceived trustworthiness and credibility when promoting sponsored products, and subsequently affect followers' brand awareness and purchase intentions (Lou and Yuan 2019). Given that influencers risk being perceived as "sellouts" if they endorse commercial products (Kozinets et al. 2010), it is imperative to understand how vocal characteristics can be leveraged to enhance the perceived authenticity and sincerity of influencers' sponsored posts. Our findings offer important insights for marketers seeking to maximize the effectiveness of influencer marketing campaigns. Our research is particularly relevant in light of recent surveys, such as Bazaarvoice's 2018 survey of 4,000 consumers, which found that a majority of consumers (62%) believe that influencer endorsements take advantage of impressionable audiences.[1]

We contribute to the literature on influencer marketing by examining influencers' vocal strategies to mitigate the negative impact of sponsorship on consumer sentiment. When influencers accept sponsorships, they may face negative consumer reactions and backlash on social media platforms, which can harm their credibility and trustworthiness (Lawson 2021). However, the specific vocal strategies that influencers can employ to improve consumer sentiment toward sponsored videos have not been well-explored. We fill this gap by investigating the effects of five primary vocal characteristics on consumer sentiment and identifying the vocal strategies that are effective in this context. In addition, we utilize several state-of-the-art deep learning models, including GAN, Mask R-CNN, Attractiveness CNN, BERTopic, and DeepFace, to analyze influencer data in novel ways. By doing so, we extend the

[1] https://www.bazaarvoice.com/press/content-called-out-47-of-consumers-fatigued-by-repetitive-influencers/

frontiers of influencer marketing research and demonstrate the value of incorporating advanced data analysis techniques in this field.

DATA

Sample Construction, Treatment, and Outcome

In our research, we utilized third-party websites that specialize in gathering information on social media influencers to identify the top 10,000 female influencers in the how-to & style category on a major video-sharing social media platform, based on their follower count. We restricted our sample to only include female beauty influencers to ensure that any differential reactions of consumers to vocal changes in sponsored videos are not confounded by gender.[2] We collected the titles and descriptions of the 30 most recent videos for each influencer, resulting in a total of 300,000 videos. We then utilized text-based classification to categorize the influencers into six subcategories: beauty, cooking, DIY, fitness, kids, and other. Our decision to focus on the beauty subcategory was motivated by the fact that it had the highest volume of brand sponsorship between 2009 and 2017, according to Schwemmer et al. (2018).

We conducted an extensive analysis of video metadata from 1,281 beauty influencers who had over 50,000 followers as of February 19, 2019. This analysis included all videos that these influencers uploaded from the beginning of their account through the end of 2018, comprising a total of 388,822 videos. The metadata of each video included various attributes such as the video ID, uploader ID, channel ID, video URL, upload date, video title, video

[2] A survey by Statista Research Department shows that 84% of influencers creating sponsored posts on Instagram were female (Statista Search Department, 2021).

description, category, tags, video duration, view count, like count, dislike count, and thumbnail images. By analyzing this metadata, we identified which of the 1,281 beauty influencers had at least one sponsored video between January 1, 2016, and December 31, 2018, resulting in 1,079 eligible influencers for further analysis.

The focus of this research is on determining whether a video is ***sponsored or not***, with the sponsored status being the treatment of interest. To identify sponsored videos, we first check for a video sponsorship disclaimer, as required by the Federal Trade Commission (FTC). However, it is possible for some videos to have undisclosed sponsorships. To ensure the robustness of our analysis, we conduct four case analyses that verify that the inclusion of the undisclosed sponsorships does not alter our results. Please refer to Web Appendices 1.A to 1.D for a detailed description of these analyses.

Consumer sentiment is a key outcome variable in our research due to its predictive power of brand performance outcomes, such as sales conversion (Schneider and Gupta 2016). Consumer sentiment is widely used in marketing research and is measured based on the text of the comments on the video. We utilize the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool (Hutto and Gilbert 2014) to calculate the sentiment of each comment on a scale of -1 to 1 and average across all comments on the video. To provide an overview of the sponsorship deal status and consumer sentiment, Table 1 presents the summary statistics based on all the 1,017 influencers included in the IV-PSM sample, which will be briefly introduced in the next paragraph and detailed in the Model section.

Table 1. Summary Statistics of Sponsorship Deal Status, and Consumer Sentiment

Sponsorship (by influencer)	N	Mean	SD	Min.	Max.
Number of all videos	1,017	102.43	92.67	2.00	1,094.00
Number of sponsored videos	1,017	16.6	16.82	1.00	146.00

Consumer Sentiment (by video)					
Non-sponsored video sentiment	16,460	0.5796	0.148	-0.9412	0.999
Sponsored video sentiment	16,460	0.5788	0.1484	-0.8639	0.9974
Notes. The counts of both sponsored and non-sponsored videos are obtained from the unmatched data, whereas the consumer sentiments are based on the data that has undergone propensity score matching.					

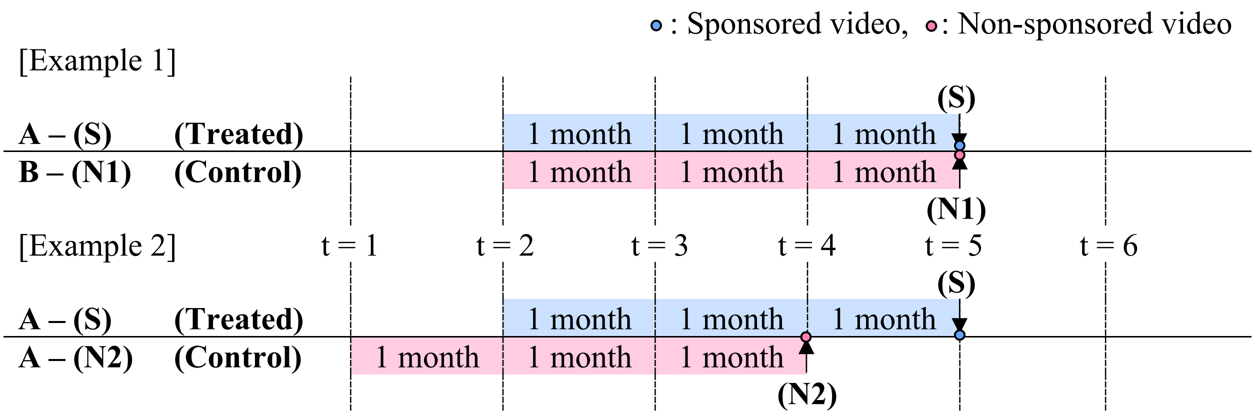
Three identification strategies are employed to estimate the effects of sponsorship and vocal characteristics (as defined in the following section) on consumer sentiment. In our IV-PSM approach, we utilized a matching technique proposed by Imai et al. (2021) to create an equivalent number of treated (sponsored) and control (non-sponsored) videos. The matching model developed by Imai et al. (2021) constructs matched control units based on (1) identical or nearest treatment history, and (2) similar characteristics to each treated unit. Given the applicability of their matching model to our influencer (non-) sponsored video data, where units may switch treatment status multiple times over time,[3] we adopt the Imai et al. (2021) model.

In matching a focal treatment (sponsored) video with a suitable control (non-sponsored) video, there are two possible sources for the control video: either it comes from a different influencer or from the same influencer's own history. To clarify this process, we present two examples in Figure 1 involving two influencers (A and B) and three videos (S, N1, and N2). At time t=5, influencer A posted a sponsored video (S), while influencer B posted a non-sponsored video (N1) at the same time. Using PSM and pre-treatment observables such as the number of previously sponsored videos and video aesthetics, (S) and (N1) can be matched. In Example 2, we consider a different scenario where influencer A posted a non-sponsored video (N2) at t=4 that can be matched with her own sponsored video (S) at t=5. This matching method allows us to account for the likelihood that an influencer may upload both sponsored and non-sponsored

[3] This is because it is common for influencers to upload both sponsored and non-sponsored videos, even after posting their first sponsored videos.

videos even after their initial sponsored video. We successfully obtained 16,460 matched pairs of videos, resulting in a total of 32,920 videos (16,460 treated (sponsored) videos uploaded by 1,015 influencers and 16,460 control (non-sponsored) videos uploaded by 1,006 influencers). Our final matched dataset includes a total of 1,017 unique influencers, consisting of 1,004 influencers with both sponsored and non-sponsored videos in the matched pairs, 11 influencers with only sponsored videos, and 2 influencers with only non-sponsored videos.

Figure 1. Illustration of Matching Treated Video (S) with Control Videos (N1, N2)



The matched sample in our study consists of 32,920 videos uploaded by 1,017 beauty influencers, comprising 16,460 sponsored videos (treatment group) and 16,460 non-sponsored videos (control group). Our sample accounts for 92.62% of the original population of 1,079 beauty influencers who had sponsorships and over 50,000 followers. The 1,017 influencers in our sample uploaded a total of 104,167 videos during the sample period, representing 93.76% of the 111,100 videos uploaded by the 1,079 beauty influencers. Notably, among the 104,167 videos, 16,886 are sponsored, which constitutes 99.4% of the 16,987 sponsored videos uploaded by the 1,079 beauty influencers. Moreover, our PSM model is capable of handling 16,460 sponsored videos, accounting for 96.9% of the 16,987 sponsored videos in the original population.

In the FRD approach, we selected the top ten cosmetic brands with the most sponsorships in our sample and obtained a sample of 589 eligible influencers who were sponsored by these brands. Out of these 2,514 videos, 1,257 were sponsored while 1,257 were non-sponsored, taking into account that some influencers had multiple videos sponsored by the top ten cosmetic brands. We used two consecutive videos from each of the 589 influencers, one before and one after the sponsorship threshold, for our analysis using the FRD approach.

Finally, for the two-way fixed effects estimators (TWFE) approach, we employ all videos uploaded by the 1,079 beauty influencers with sponsorships and at least 50,000 followers.

The potential for selection bias in our effect estimation arises if there exist differences between consumers who view sponsored and non-sponsored videos. To mitigate this risk, we limit our analysis to comments from consumers who have commented on at least one sponsored and one non-sponsored video in our dataset. By applying this inclusion criteria, we are left with a total of 32,920 videos containing 4,957,744 comments, which accounts for 57.2% of all 8,669,021 comments.

Vocal Characteristics

In accordance with prior research on vocal characteristics (Tusing and Dillard, 2000; Fraccaro et al. 2011; Zuckerman and Driver 1989) and literature on affective computing (Scherer et al. 1973) and ASR (Eyben et al. 2010; Schuller et al. 2019), we include five vocal characteristics in our analysis. These vocal characteristics are the average loudness, average pitch, loudness variability, pitch variability, and talking duration, which we extract with OpenSMILE (Eyben et al. 2010; Schuller et al. 2019). OpenSMILE is a widely-used, open-source toolkit for extracting acoustic characteristics in the affective computing and ASR literature (Chakraborty et al. 2015). We provide a brief summary of each characteristic, and details about voice feature extraction appear in Web Appendices 2.A and 2.B. The summary statistics presented in this section are based on the 1,017 influencers in the IV-PSM sample, consistent with those in Table 1.

The vocal characteristic of ***average loudness***, which is measured by sound intensity, is a narrow band approximation of the voice signal and typically represented in decibels (dB). While consumers have the ability to control the volume of the videos they watch, it is reasonable to assume that they maintain the same volume setting when watching both sponsored and non-sponsored videos. Consequently, if influencers alter their average loudness considerably in sponsored videos as compared to non-sponsored videos, it is likely that consumers will perceive a difference in loudness.

$$Vocal\ intensity\ (dB) = 10 \times \log_{10} \frac{Vocal\ intensity\ (Watts/meter^2)}{I_0}$$

where the reference intensity $I_0 = 1 \times 10^{-12} W/m^2$

Existing research on the effect of loudness on consumer behavior in the business context is inconclusive. While higher loudness was found to be associated with lower refusal rates in telephone survey providers (Oksenberg et al. 1986), a research on banking telemarketing showed that moderate loudness was perceived as more credible than high loudness (Chebat et al. 2007).

The vocal characteristic ***average pitch*** is measured by F0, the fundamental frequency of the voice signal, which is expressed in units of Hz. The typical adult female's pitch range is between 165 to 255 Hz, while the typical adult male's range is between 85 to 155 Hz. The literature has yielded mixed results concerning the effects of average pitch on social outcomes.

For instance, higher pitch has been associated with higher attractiveness in mate selection (Fraccaro et al. 2011). However, in other contexts, lower-pitched voices were found to be more trustworthy for leadership roles (Klofstad et al. 2012), and lower-pitched voices have been associated with election success (Banai et al. 2017).

Loudness variability is measured by the zero-crossing rate, which is the number of times the amplitude of the speech signal passes through zero in a given time interval. Existing research on loudness variability in the business context is scarce. However, a study found that speakers with higher loudness variability were more effective at maintaining attention and perceived as more attractive (Zuckerman and Driver 1989).

Pitch variability is defined as the standard deviation of the fundamental frequency (F0) (Banai et al. 2017). Prior research in the business context has yielded mixed findings. Specifically, higher pitch variability was found to result in lower refusal rates in telephone surveys (Oksenberg et al. 1986). In contrast, in the context of political elections, candidates with lower pitch variability tended to emerge as winners (Banai et al. 2017).

Talking duration is an acoustic feature that measures the amount of time an influencer spends speaking in a video. It is determined by the voicing probability, which is computed by measuring the inverse of the maximum frequency via the autocorrelation function. In essence, the voicing probability captures the percentage of time during which the influencer is speaking in the video. Notably, a higher talking duration implies that the influencer talks more during the video. In a relevant literature in the business domain, Peterson et al. (1995) found that salespeople who talked for shorter durations had higher sales output.

We examine five vocal characteristics and their variation in magnitude, namely average loudness, average pitch, loudness variability, pitch variability, and talking duration. To compare the relative change between sponsored and non-sponsored videos, the raw values of these characteristics are standardized across individuals. Table 2 provides a summary of the standardization process and the resulting summary statistics of the vocal characteristics. It should be noted that the standardization process does not alter the relative change between the sponsored and non-sponsored videos, but rather facilitates the comparison of the magnitude of the vocal characteristics across the videos.

Table 2. Summary Statistics of Vocal Characteristics

Vocal Characteristics (by video)	N	Mean	SD	Min.	Max.
Average loudness (intensity in W/m²)[4]	32,920	$2.95\text{x}10^6$	$3.86\text{x}10^6$	0.00	$9.34\text{x}10^5$
Average pitch (F0 in Hz)	32,920	187.63	33.8	0.2	337.5
Loudness variability (zero-crossing rate)	32,920	0.07	0.03	0.01	0.99
Pitch variability (standard dev. Of F0)	32,920	148.14	18.13	7.96	189.23
Talking duration (voicing probability)	32,920	0.61	0.04	0.00	0.83

Control Variables

Our research employs the 3Vs model, which encompasses vocal, visual, and verbal characteristics to characterize video content. The significance of these features has been established in prior research and theories (Mobius and Rosenblat, 2006; Li et al., 2018; Schwanenflugel et al., 1988), and we use them to identify their impact empirically. Moreover, we utilize metadata to incorporate measures of popularity. While physical methodologies such as sound pressure are used to measure vocal characteristics, human ratings collected from the

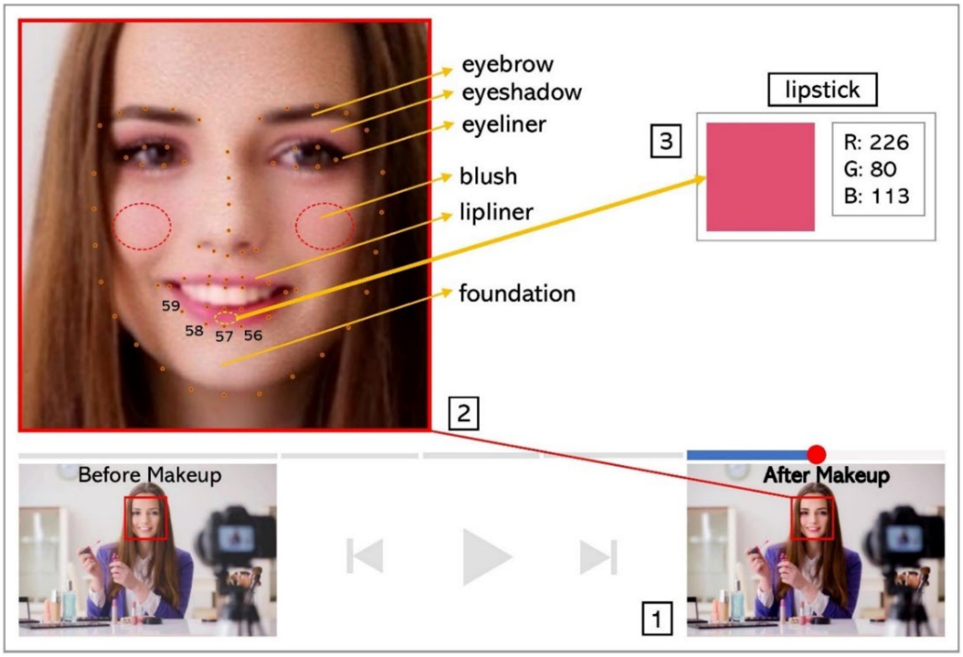
[4] Loudness in Watts/meter^2 can be converted into decibels (dB), a more familiar unit of loudness. The mean loudness in dB is 64.69, the standard deviation is 65.87 dB, and the min & max values are 25.18 dB and 79.70 dB, respectively.

experiment dataset are used to assess other features, such as facial attractiveness. As these visual and verbal features are subjective, we view them as controls and exercise caution in our interpretation. We refrain from overinterpretation, such as exploring the relationship between influencer attractiveness and consumer sentiment, and focus on the key variables (i.e., voice features) in our analysis.

Visual characteristics. We incorporate six distinct sets of visual characteristics in order to assess their impact on consumer sentiment. The sets include makeup colors, makeup spectrum (i.e., the heaviness of the makeup), Neural Image Assessment (NIMA) score (Talebi and Milanfar, 2018), facial attractiveness, demographics and emotions, and objects. Two makeup characteristics, namely makeup colors and spectrum, were included as a majority of videos in our sample featured influencers applying makeup products. The remaining four sets were included due to their potential impact on the overall aesthetic appeal of the influencer and the video.

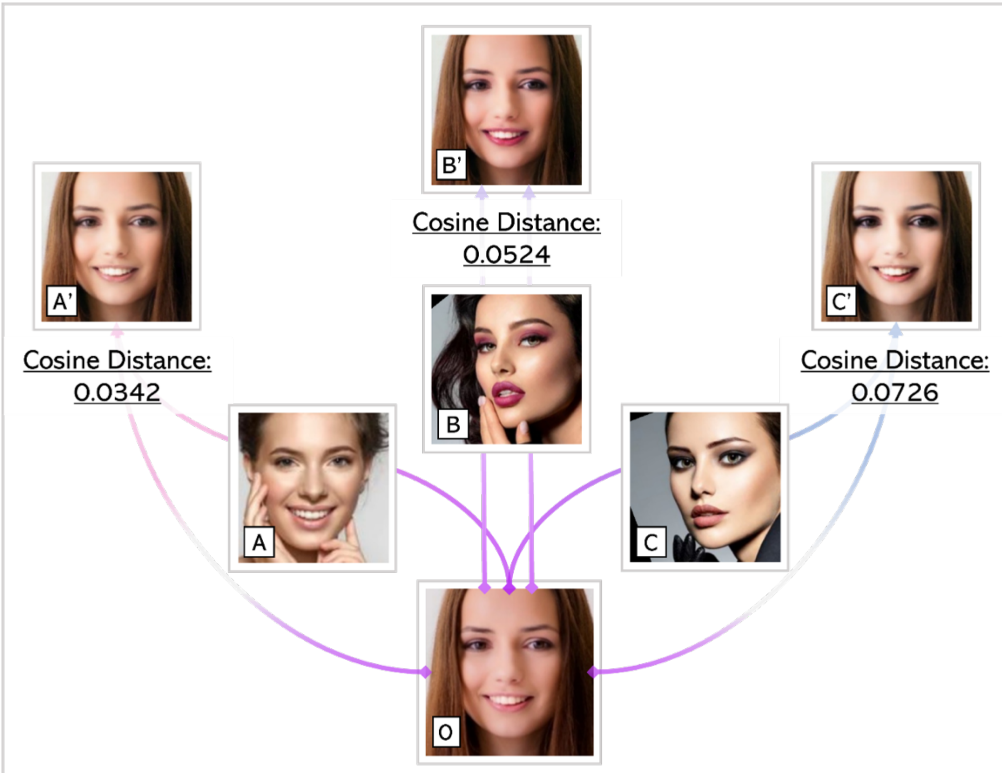
The potential impact of differing makeup colors on consumer sentiment between sponsored and non-sponsored videos is an important consideration when estimating treatment effects. To address this, we utilize Dlib, a pre-trained model in the Python library, to detect the G (Green) values in 17 facial areas where makeup is applied. These facial areas are identified using 68 facial landmarks, and their RGB values are extracted. By controlling for makeup colors in this way, we can reduce the possibility of biasing our estimated treatment effects due to differences in aesthetics. Figure 2 illustrates the 17 facial landmarks used in this analysis, and further details can be found in Web Appendix 2.C.

Figure 2. Makeup-Applied Facial Landmarks for Makeup Color Measurements



In the context of sponsored and non-sponsored videos, differences in ***makeup heaviness*** can affect consumer sentiment and potentially bias estimations. To control for this, we utilize BeautyGAN, a Beauty Generative Adversarial Network model introduced by Li et al. (2018), to quantify the heaviness of makeup application. The model enables the precise transfer of makeup style from a reference makeup face image to a source image. By utilizing BeautyGAN, we can measure the degree of heaviness of makeup applied in sponsored and non-sponsored videos, thus accounting for potential bias in estimation.

Figure 3. Examples of Reference Video Images (A, B, and C) and Makeup-Transferred Video Images (A’, B’, C’) with Cosine Distance Scores



To quantify the heaviness of the makeup applied by influencers, we employ a two-step process, elaborated in Web Appendix 2.D. In the first step, we use the Beauty Generative Adversarial Network (BeautyGAN) model to transfer the makeup style from an exemplar facial image to an influencer’s baseline facial image. In the second step, we compute the cosine distance between the baseline image and the makeup-transferred image, where a higher cosine distance score suggests a more significant change in the baseline image and hence a relatively heavier makeup application. To illustrate, Figure 3 presents an example of an influencer’s baseline image (bottom center), three exemplar images from our video dataset (A, B, and C), and their corresponding makeup-transferred images (A’, B’, and C’) with associated cosine distance scores. The cosine distance scores reveal that the influencer in image A applied the lightest makeup, whereas the influencer in image C used the heaviest makeup. We refer interested readers to Web Appendix 2.D for a detailed description of the process.

The evaluation of ***visual aesthetics*** associated with influencer content plays a pivotal role in determining the impact of the content on consumer sentiment. In this regard, the Neural Image Assessment (NIMA) score, a predictive model developed by Talebi and Milanfar (2018), is leveraged to assign a quality score to influencer videos. To mitigate computational complexity concerns, we select the MobileNet model from the three models proposed by Talebi and Milanfar (2018) due to its computational efficiency. Further details are provided in Web Appendix 2.E.

In the context of influencer marketing, higher levels of ***facial attractiveness*** have been found to elicit more positive responses from consumers (Mobius and Rosenblat, 2006). Therefore, it is important to account for potential bias in estimation due to variations in influencer attractiveness between sponsored and non-sponsored videos. To address this issue, we train a convolutional neural network (CNN) model on the SCUT-FBP5500 dataset (Liang et al., 2018) to score each influencer's attractiveness on a scale of 1 to 5. The detailed methodology for the attractiveness scoring process can be found in Web Appendix 2.F.

To examine the possibility of confounding effects stemming from ***demographic and emotional characteristics***, which may impact consumer sentiment through their compatibility with the sponsoring brand or product, we adopt the DeepFace framework developed by Taigman et al. (2014) from the Facebook research team. This deep learning-based system enables us to classify an influencer's race into one of six categories (Asian, Black, Indian, Latino/Hispanic, Middle Eastern, or Non-Hispanic White) and to evaluate the influencer's emotional state in terms of seven emotions (angry, disgust, fear, happy, neutral, sad, and surprise). While acknowledging the unpredictability of demographic and emotional influences, we contend that such factors could have a bearing on consumer sentiment, hence the relevance of our analysis. For a more comprehensive summary, please consult Web Appendix 2.G.

We address the potential influence of ***objects*** worn or displayed by influencers in sponsored versus non-sponsored videos on consumer sentiment. Utilizing the iMaterialist fine-grained visual segmentation training dataset, we train the Mask R-CNN model (He et al. 2017) to detect and classify 17 apparel objects commonly seen in videos: top/t-shirt/sweatshirt, jacket, pants, skirt, dress, glasses, hat, watch, belt, tights/stockings, shoe, bag/wallet, collar, lapel, sleeve, pocket, and neckline. Figure 4 demonstrates an example of our object detection results, while Table 5 presents summary statistics of the visual characteristics. For detailed information regarding all objects, please refer to Web Appendix 2.H.

Figure 4. Examples of Mask R-CNN Object Detection Results

Verbal characteristics. We employ video transcripts to extract three distinct verbal characteristics, namely narrative topics, verbal sophistication, and concreteness. We hypothesize that these attributes might display systematic differences between sponsored and non-sponsored videos as the sponsoring brand typically develops or reviews the script for sponsored videos.

To address potential bias from systematic differences in ***narrative topics*** between sponsored and non-sponsored videos, we use the BERTopic model (Grootendorst 2020), a state-of-the-art topic modeling technique that utilizes BERT transformers and class-based TF-IDF to generate dense clusters. Through Uniform Manifold Approximation and Projection (UMAP), BERTopic reduces the dimension of BERT embeddings and generates topic clusters automatically. We extract nine verbal narrative topics, including hair, lip, crease, fashion, skincare, applying, palette, brow, and brush, using Table 3 to display the top 5 keywords with the highest probabilities. For a more detailed explanation, please refer to Web Appendix 2.I.

Table 3. Verbal Narrative Topics and Top 5 Keywords

Topic	Words
Hair	Hair, curls, wig, conditioner, section
Lip	Lip, lipstick, color, liquid, nude
Crease	Crease, brush, using, going, shade
Fashion	Dress, wear, cute, jeans, jacket
Skincare	Skin, skincare, acne, face, cleanser
Applying	Using, going, brush, apply, concealer
Palette	Palette, shade, shadow, color, eyeshadow
Brow	Brow, eyebrow, brush, hair, pencil
Brush	Brush, use, one, sigma, foundation

We employ three lexical measures of ***verbal sophistication***: unigram, bigram, and trigram frequency in the Corpus of Contemporary American English (COCA). Infrequently used words in COCA, such as edifice, cuisine, and egregious, are considered more sophisticated than frequently used words like building, food, and bad. We calculate a score for verbal sophistication based on the frequency of these words in the transcript. Table 4 presents examples of high and low verbal sophistication sentences describing makeup foundation. For further details, see Web Appendix 2.J.

Table 4. Examples of Sentences with High versus Low Sophistication Scores

Sentence	Sophistication (COCA Corpus)
"…….This foundation provides exceptional coverage without compromising a smooth striking finish formulated with 21 percent pigment this powerful formula was developed with proprietary pigments ……"	1,447.50
"……I feel like my skin just honestly looks a little bit dull currently with this foundation so I would say not for like normal to dry skin more for like combo oily skin types I can totally see why this is popular ……"	1,403.61

We examine the impact of ***verbal concreteness*** on consumer sentiment by using Brysbaert's Concreteness measure, which includes 37,058 English words and 2,896 bigrams, based on ratings from over 4,000 crowdsourced participants in a norming study (Brysbaert et al. 2014). We expect that concrete words will be more memorable to listeners, and therefore, have an effect on consumer sentiment. Table 5 provides summary statistics of select verbal characteristics.

Additional controls from metadata. The analysis also incorporates several metadata variables, in addition to the 3V characteristics.

We incorporate ***influencer popularity*** as a control variable, as it has the potential to impact sponsorship opportunities and consumer sentiment. We mitigate for short-term popularity fluctuations (e.g., a popularity decreases due to a publicized scandal) by examining daily changes in follower counts, total views, and number of Instagram followers (gathered from Facebook's

CrowdTangle), while we address long-term popularity by examining the raw counts of these same measures.

To control for potential confounding effects of a brand's concurrent marketing efforts on consumer sentiment toward sponsored videos, we include the brand's monthly ***advertising spending*** as a covariate in our analysis. We use data from Kantar Media's Ad$pender database, which tracks advertising expenditures and occurrences on 18 media platforms for over 3 million brands, to account for unknown shocks in the brand's promotional activities. In addition, we conduct a robustness check by including media-specific spending amounts.[5]

In the final model, we also control for fixed effects of the 30 most frequently featured brands, the 30 most frequently featured product types, and eight video-specific characteristics (including video duration, view count, like count, dislike count, comment count, title length, description length, and the influencer's tenure). Additionally, we include six video format dummy variables to account for video content type. Summary statistics of video characteristics, popularity, and advertisement spending are presented in Table 5, with the complete report available in Web Appendices 2.K, 2.L, and 2.M.

Table 5. Summary statistics of Selected Visual/Verbal/Video Characteristics & Ad Spending

Visual Characteristics (by video)	N	Mean	SD	Min.	Max.
Eyebrow color G value	32,920	0.38	0.2	0.00	1.00
Makeup heaviness cosine dist.	32,920	0.06	0.04	0.00	0.28
Image aesthetics (Semi-log NIMA)	32,920	1.72	0.5	0.00	2.05
Demographics – Race:White (%)	32,920	0.34	0.32	0.00	1.00
Emotion – Neutral (%)	32,920	0.33	0.36	0.00	1.00
Object – Sleeve	32,920	0.66	0.47	0.00	1.00
Verbal Characteristics (by video)					

[5] The 18 media platforms covered by Ad$pender data are Network TV, Spot TV, Spanish Language Network TV, Cable TV, Syndication, Magazines, Sunday Magazines, Local Magazines, Hispanic Magazines, B-to-B Magazines, National Newspapers, Newspapers, Hispanic Newspapers, Network Radio, National Spot Radio, Local Radio, US Internet, and Outdoor.

Topic: Hair	32,920	0.11	0.31	0.00	1.00
Sophistication: Unigram (Semi-log)	32,920	6.27	2.44	0.00	8.25
Video Characteristics (by video)					
Duration (Semi-log)	32,920	6.36	0.58	2.64	9.16
Like count (Semi-log)	32,920	7.17	1.43	1.95	12.63
Tenure (Semi-log)	32,920	7.15	0.68	0.00	8.37
Popularity & Ad Spending (by video)					
Follower count (Semi-log)	32,920	12.26	1.19	6.69	15.11
Monthly Ad. Spending (Semi-log)	32,920	3.97	3.71	0.00	10.56
Notes. For visual characteristics, we provide the statistics for only the most common race, emotion, and object. For the full summary statistics, see Web Appendix 2.					

MODEL

This section presents the model specifications. First, we elaborate on the application of PSM to identify whether influencers alter any of their vocal characteristics in sponsored videos in comparison to non-sponsored ones. Then, we introduce three techniques (IV, FRD, and TWFE) to estimate the impact of sponsorship, vocal characteristics, and their interplay on consumer sentiment. Finally, we illustrate the approach for characterizing individual variability in the utilization of the voice modulation technique.

Modulation of Vocal Characteristics in Sponsored (vs. Non-Sponsored) Videos: Propensity Score Matching

Our primary objective is to estimate the impact of sponsorship on voice by drawing comparisons between sponsored and non-sponsored videos that have comparable pre-treatment history, influencer attributes, and video characteristics. Nevertheless, we encounter three modeling challenges when matching influencer-video pairs.

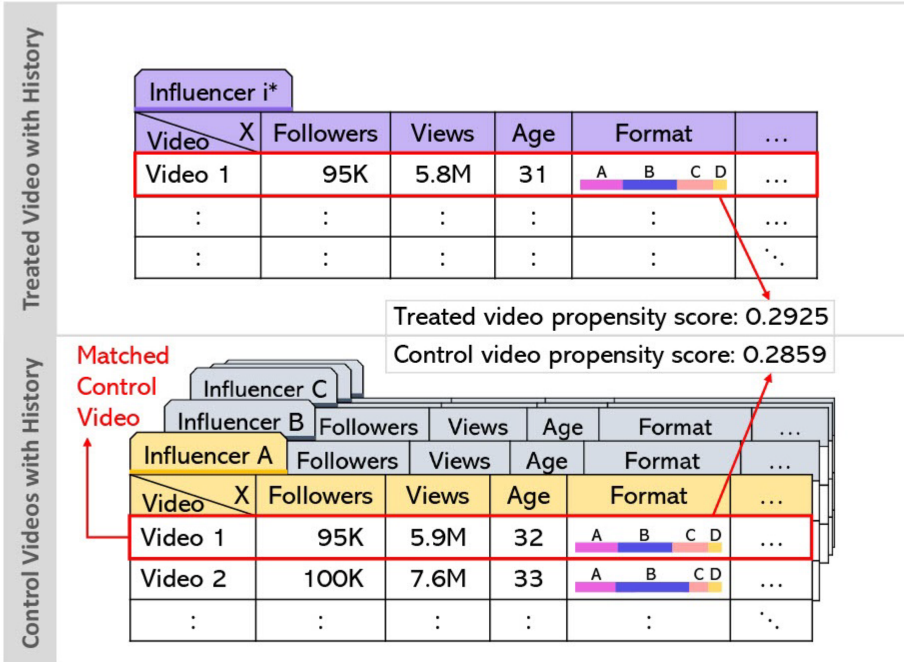
Table 6. Propensity Score Matching (PSM) Modeling Challenges & Solutions

No.	Modeling Challenge	Solution

1.	Firstly, brands tend to choose influencers based on their previous video history, rather than a single video.	We construct quarterly sponsorship history, as well as influencer and video characteristics for each video, which are utilized as pre-treatment observables in PSM.
2.	The timing of sponsorship treatment may vary across influencers.	We control for influencer and time-fixed effects and conduct matching at the video level.
3.	An influencer can post non-sponsored videos even after posting the first sponsored (treated) video.	We employ the identification strategy presented by Imai et al. (2021) to accommodate treatment reversal.

Following the approach described in Table 6, we obtain 16,460 matched pairs of sponsored and non-sponsored videos. These pairs are constructed to be comparable on pre-treatment quarterly observable history and characteristics, including sponsorship history, 3V characteristics (vocal, visual, and verbal), textual features, and influencer popularity. The matching process is illustrated in Figure 5, where treated (sponsored) and control (non-sponsored) videos are shown. Specifically, the upper box displays a treated video observation of influencer i* (Video 1) with her pre-treatment quarterly history and characteristics, while the lower box shows control video observations of influencer A, B, C, and others with their pre-treatment quarterly history and characteristics. Influencer A's Video 1 (non-sponsored video) is selected to be the matched control video of influencer i*'s Video 1 (sponsored video) based on pre-treated quarterly observables, such as follower count, view count, age, and video format distribution.

Figure 5. Illustration of Propensity-Score-Matched Control Video Selection Process



To calculate propensity scores for the PSM analysis, we employ a logistic regression model (Equation 1) that predicts whether influencer i's video j was sponsored at a given time t. Details of the summary statistics for the PSM-matched data can be found in Web Appendix 3.A.

$$Sponsorship_{ijt} = \beta_0 + \overrightarrow{\beta_1} \times \overrightarrow{TreatmentHistory_{i,t-Q}} + \overrightarrow{\alpha_0} \times \overrightarrow{VisualFeatures_{i,t-Q}} + \overrightarrow{\alpha_1} \times \overrightarrow{VerbalFeatures_{i,t-Q}} + \overrightarrow{\alpha_2} \times \overrightarrow{TextualFeatures_{i,t-Q}} + \overrightarrow{\alpha_3} \times \overrightarrow{Popularity_{i,t-Q}} + \overrightarrow{\alpha_4} \times \overrightarrow{Brands_{i,t-Q}} + \overrightarrow{\alpha_5} \times \overrightarrow{Types_{i,t-Q}} + \alpha_6 \times AdSpending_{i,t-Q} + I_i + \tau_t + \epsilon_{ijt}, \epsilon_{ijt} \sim N(0, \sigma_\epsilon^2)$$

$$\overrightarrow{TreatmentHistory_{i,t-Q}} = [Video_{t-Q}, Sponsorship_{t-Q}, SponsorshipCategory_{t-Q}]' \quad (1)$$

$$\overrightarrow{VisualFeatures_{i,t-Q}} = [\overrightarrow{MakeupStyle_{i,t-Q}}, MakeupSpectrum_{i,t-Q}, Attractiveness_{i,t-Q}, Age_{i,t-Q}, \overrightarrow{Race_{i,t-Q}}, \overrightarrow{Objects_{i,t-Q}}]'$$

$$\overrightarrow{VerbalFeatures_{i,t-Q}} = [\overrightarrow{Topics_{i,t-Q}}, \overrightarrow{Sophistications_{i,t-Q}}]'$$

$$\overrightarrow{TextualFeatures_{i,t-Q}} = [Duration_{i,t-Q}, Views_{i,t-Q}, Likes_{i,t-Q}, Dislikes_{i,t-Q}, Comments_{i,t-Q},$$

$$TitleLength_{i,t-Q}, DescLength_{i,t-Q}, Tenure_{i,t-Q}, \overrightarrow{Formats_{i,t-Q}}]$$

$$\overrightarrow{Popularity_{i,t-Q}} = [Followers_{i,t-Q}, AccountViews_{i,t-Q}, InstagramFollowers_{i,t-Q}, \Delta Followers_{i,t-Q}, \Delta AccountViews_{i,t-Q}, \Delta InstagramFollowers_{i,t-Q}]'$$

The analysis is conducted at the level of the influencer-time observation, where the dependent variable $Sponsorship_{ijt}$ takes a value of 1 if influencer *i*'s video j is sponsored at time *t*. The vector $\overrightarrow{TreatmentHistory_{i,t-Q}}$ consists of three elements, namely, the number of videos ($Video_{t-Q}$), the number of sponsored videos ($Sponsorship_{t-Q}$), and the number of sponsored videos by category ($SponsorshipCategory_{t-Q}$) posted from a quarter ago (t-Q) to time t. The aforementioned features enable the analysis of the impact of sponsorship history on the likelihood of sponsorship at time t, accounting for the temporal dimension of the influencer's content output and sponsorship activity. The vector $\overrightarrow{VisualFeatures_{i.t-\mathrm{Q}}}$ contains the G colors from the 17 makeup-applied areas ($\overrightarrow{MakeupStyle_{i,t-\mathrm{Q}}}$; from 0 to 1), cosine distance score of a makeup-transferred face ($MakeupSpectrum_{i,t-\mathrm{Q}}$; from 0 to 0.25), attractiveness score ($Attractiveness_{i,t-\mathrm{Q}}$; from 1 to 5), log-transformed age ($Age_{i,t-\mathrm{Q}}$), five races ($\overrightarrow{Race_{i,t-\mathrm{Q}}}$; probabilities from 0 to 1), and 17 object dummy variables ($\overrightarrow{Objects_{i,t-\mathrm{Q}}}$; probabilities from 0 to 1). The vector $\overrightarrow{VerbalFeatures_{i,t-\mathrm{Q}}}$ comprises of the top ten narrative topics ($\overrightarrow{Topics_{i,t-\mathrm{Q}}}$; probabilities from 0 to 1) and four verbal sophistication measures ($\overrightarrow{Sophistications_{i,t-\mathrm{Q}}}$; the log-transformed verbal concreteness score and log-transformed frequency of spoken words, bigrams, and trigrams via COCA). The vector $\overrightarrow{TextualFeatures_{i,t-\mathrm{Q}}}$ contains a vector of six video formats ($\overrightarrow{Format_{i,t-\mathrm{Q}}}$; probabilities from 0 to 1) and eight log-transformed textual measures: video duration in seconds ($Duration_{i,t-\mathrm{Q}}$), view count ($Views_{i,t-\mathrm{Q}}$), like count

($Likes_{i,t-\mathrm{Q}}$), dislike count ($Dislikes_{i,t-\mathrm{Q}}$), comment count ($Comments_{i,t-\mathrm{Q}}$), the length of the title and description in characters ($TitleLength_{i,t-\mathrm{Q}}$ and $DescriptionLength_{i,t-\mathrm{Q}}$), and the number of days influencer *i* had worked on the platform since the first video upload ($Tenure_{i,t-\mathrm{Q}}$). The vector $\overrightarrow{Popularity_{i,t-\mathrm{Q}}}$ consists of three log-transformed measures and three normalized measures, where the log-transformed measures control for long-term popularity: the number of followers ($Followers$), total account views ($AccountViews$), and the Instagram followers of influencer *i* ($InstagramFollowers$), while the three normalized measures control for short-term popularity shocks and are measured as the daily change in each variable ($\Delta Followers_{i,t-\mathrm{Q}},\ \Delta AccountViews_{i,t-\mathrm{Q}},\ \Delta InstagramFollowers_{i,t-\mathrm{Q}}$). $\overrightarrow{Brands_{i,t-Q}}$ and $\overrightarrow{Types_{i,t-Q}}$ represent the 30 most commonly featured brands and product types, respectively, while $AdSpending_{i,t-Q}$ denotes the amount of expenditure on advertisements. I_i is the indicator variable for influencer i, τ_t are the dummy variables for the upload year-month, and ϵ_{it} is the idiosyncratic error term. Propensity scores are obtained through a logistic regression model (Equation (1)) that predicts whether influencer *i*'s video *j* was sponsored at time *t*. Further details regarding the PSM-matched data can be found in Web Appendix 3.A, while the effectiveness of PSM is discussed in Web Appendices 3.B and 3.C.

We estimate Equation (2) on a PSM sample of 1,017 matched influencers with observable characteristics to investigate if there are differences in vocal characteristics between sponsored and non-sponsored videos.

$$Voice_{ijt}^{k} = \beta_0 + \beta_1 \times Sponsorship_{ijt} + \vec{\Gamma} \times \overrightarrow{ControlVariables_{ijt}} + I_i + \tau_t + \epsilon_{ijt}$$
$$\epsilon_{ijt} \sim N(0, \sigma_\epsilon^2) \qquad (2)$$
$$\text{where } \vec{\Gamma} \times \overrightarrow{ControlVariables_{ijt}} = \overrightarrow{\alpha_0} \times \overrightarrow{VisualFeatures_{ijt}} + \overrightarrow{\alpha_1} \times$$

$$\overrightarrow{VerbalFeatures_{ijt}} + \overrightarrow{\alpha_2} \times \overrightarrow{TextualFeatures_{ijt}} + \overrightarrow{\alpha_3} \times \overrightarrow{Popularity_{ijt}} +$$
$$\overrightarrow{\alpha_4} \times \overrightarrow{Brands_{ijt}} + \overrightarrow{\alpha_5} \times \overrightarrow{ProductTypes_{ijt}} + \overrightarrow{\alpha_6} \times \overrightarrow{Ad\ Spending_{ijt}}$$

where $Voice_{ijt}^{k}$ is vocal attribute k (k ∈{*average loudness, average pitch, loudness variability, pitch variability, talking duration*} in influencer i's video j at time t. β_0 denotes baseline vocal characteristics, and β_1 represents the treatment effect, indicating the change in vocal characteristics in sponsored videos compared to non-sponsored videos. A negative β_1 for average loudness, for instance, would imply that influencers speak more softly in sponsored videos than in non-sponsored ones.

Impact of Vocal Attributes and Sponsorship on Consumer Sentiment

Our objective is to examine the impact of vocal attributes and sponsorship on consumer sentiment. However, unobserved variables could lead to endogeneity concerns and bias the estimated treatment effect. Specifically, both sponsorship and vocal characteristics could be sources of endogeneity. For example, winning an influencer award may simultaneously increase sponsorship offers and improve consumer sentiment on sponsored posts. Additionally, significant events could impact the influencers' vocal attributes. Negative events, such as publicized scandals, are expected to produce opposite effects. To overcome this issue, we implement three identification strategies: IV, FRD, and TWFE.

IV approach.

We employ the instrumental variable (IV) method on the sample generated from propensity score matching (PSM). To instrument the endogenous sponsorship variable, we use the number of concurrent videos (brand m and parent company m') sponsored by brand m and the parent company m' in week t, which is the week in which influencer i posted video j. The instrument satisfies the relevance condition, as firms commonly engage with multiple influencers at the same time. Thus, the probability of influencer i receiving a sponsorship from brand m and parent company m' in week t should positively correlate with the number of sponsorships from brand m in week t. For example, L'Oréal League collaborated with 15 influencers (WWD, 2016), and SephoraSquad program engaged with 24 influencers (Fast Company, 2019). In the IV estimation, it is crucial that the instruments satisfy the exclusion restriction, where concurrent sponsorship decisions of firms and parent holding companies are not correlated with individual influencers' popularity shocks. We hypothesize that there exists no correlation between the consumer sentiment elicited by an influencer's videos, as gauged by the associated comments, and the number of other influencers enlisted by the sponsoring brands of the focal influencer (for supporting evidence, refer to Web Appendix 3.D and 3.E). Intuitively, the absence of a correlation is plausible because the concurrent recruitment decisions of sponsoring brands for numerous influencers are unlikely to be influenced by the idiosyncratic shocks experienced by a single influencer. Specifically, we operationalize the sponsoring brand's simultaneous sponsorship decisions as the weekly count of sponsorships by the brand.

Using two-stage least squares, the first stage equation is

$$Sponsorship_{ijt} = \beta_0 + \overrightarrow{\beta_1} \times \overrightarrow{Concurrent\ videos\ (m,m')} + \overrightarrow{\beta_2} \times \overrightarrow{Voice_{ijt}^{k}} \times \overrightarrow{Concurrent\ videos\ (m,m')} + \overrightarrow{\Gamma} \times \overrightarrow{ControlVariables} + I_i + \tau_t + \epsilon_{ijt} \tag{3}$$

Next, we utilize the sponsorship instrument ($\widetilde{Sponsorship}_{ijt}$) and the interaction instrument between sponsorship and vocal attributes ($\widetilde{Sponsorship}_{ijt} \times \overrightarrow{Voice_{ijt}^{k}}$) to derive estimates for the second-stage equation:

$$Sentiment_{ijt} = \beta_0 + \beta_1 \times \widetilde{Sponsorship}_{ijt} + \beta_2 \times \overrightarrow{Voice_{ijt}^{k}} + \beta_3 \times \tag{4}$$

$$\widetilde{Sponsorship_{ijt}} \times \overrightarrow{Voice_{ijt}^{k}} + \vec{\Gamma} \times \overrightarrow{ControlVariables} + I_i + \tau_t + \epsilon_{ijt},$$
$$\epsilon_{ijt} \sim N(0, \sigma_\epsilon^2)$$

For the purpose of robustness checks, we further conduct estimations for Equation (3) and (4) using instruments for both the vocal characteristics and the interaction between sponsorship and vocal characteristics. We present the outcomes of these analyses in the Online Appendix 6.A for reference.

FRD approach.
Fuzzy regression discontinuity (FRD) is a design that extends regression discontinuity (RD) by estimating treatment effects when treatment is based on an observed assignment variable (Lee and Lemieux 2010).

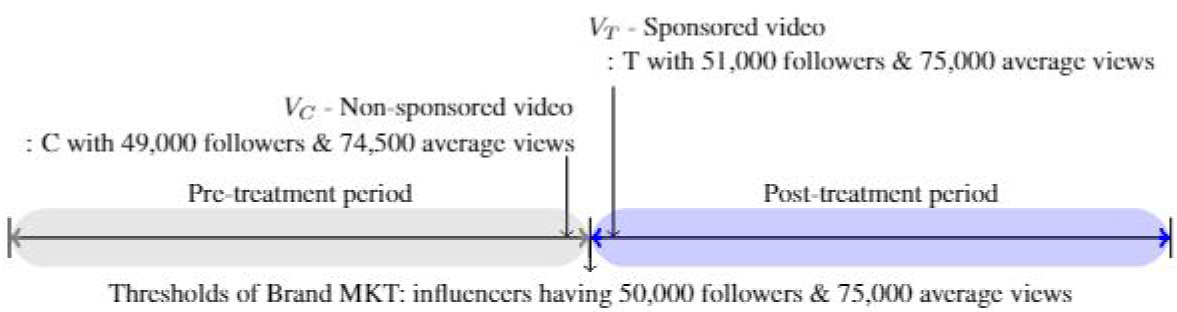
In a qualitative interview with Aaron Layden,[6] a sales manager at Neoreach, an influencer campaign consulting firm, it was revealed that follower count and average views are two commonly used popularity metrics that brands consider when selecting influencers for their campaigns. This is corroborated by evidence gathered from influencer application forms, email responses from marketing managers, and executive interviews, which can be found in Web Appendix 4.A. In light of this information, we adopt both follower count and average views as assignment variables and utilize a regression discontinuity design that accounts for multiple assignment variables, following the approach proposed by Papay et al. (2011).

Figure 6 is provided as an illustration of the treatment timeline and rationale behind the RD design. Suppose influencer *i* exists both before and after the treatment period. Influencer *i* becomes a beneficiary of a sponsorship (i.e., treatment) from a hypothetical brand MKT after

[6] We conducted a phone interview on September 20, 2019, and an email interview that ended on October 1, 2019.

exceeding follower count and average view thresholds of 50,000 and 75,000, respectively. (For simplicity, we assume that all influencers who surpass these thresholds are sponsored by MKT.) The focus is on influencer $\underline{i}$'s first video (V_T, the subscript T denotes treatment) after the treatment, and the last video (V_C, the subscript C denotes control) before the treatment. We compare the sponsored video (V_T) with the non-sponsored video (V_C) to estimate the impact of sponsorship, vocal characteristics, and their interactions.

Figure 6. An Example Treatment Timeline in the FRD Approach



To address potential issues of recurring sponsorships being influenced by previous performance of the influencer, we limit our analysis to only the first sponsored video in a brand-influencer pair. Previous research has suggested that sponsorship decisions may be influenced by factors beyond the popularity metrics we consider (Barker 2018), leading to potential weak instrument problems and violations of the RD assumption (Lee and Lemieux 2010). Thus, we employ a selective approach in which we only consider the first sponsored video for each brand-influencer pair. Further details on our approach can be found in Web Appendix 4.B.

We restrict our analysis to the top ten brands that hired the most influencers (as reported in Table 7), as smaller brands with fewer sponsorships may have additional criteria for selecting influencers, potentially violating the fundamental RD assumption.

We employed three methods to validate that the top ten brands hire influencers based on the two assignment variables (follower count and average views). Firstly, we conducted email interviews with the Marketing/PR managers of each brand to understand their influencer selection process. Secondly, we reviewed brand-generated articles that provide insights into the sponsorship selection process. Thirdly, we examined the influencer applications to ensure that the brands requested information on the two assignment variables. Additional information can be found in Web Appendix 4.A. Table 7 provides a list of the top brands and their corresponding thresholds for minimum follower count and minimum average views of the sponsored influencers. The representativeness of the ten brands is described in Web Appendix 4.C.

Table 7. Follower Count and Average View Thresholds of the Top 10 Brands

Brand	Follower Count Threshold	Average Views Threshold
Fabfitfun	16,186	482
Garnier	19,492	1,677
L'Oréal Paris	27,624	2,381
Maybelline New York	17,756	1,192
Neutrogena	22,405	793
Olay	18,989	1,591
Scentbird	11,362	662
Sephora	12,931	824
Ulta	42,281	897
Vanity Planet	7,114	995

Some influencers may meet a brand's follower count and average view thresholds but may not post any sponsored videos due to various reasons, such as taking a break from sponsored content or the brand's limited marketing budget. Therefore, the presence of such unsponsored influencers above the thresholds can create "fuzziness" in the discontinuities, violating the RD assumption. To address this, we adopt the Fuzzy RD approach, as proposed by

Lee and Lemieux (2010) for modeling multiple assignment variables in the presence of fuzziness.

To comprehend the effects of sponsorship, vocal characteristics, and their interrelationships on consumer sentiment,[7] we perform a two-stage model estimation:

$$\begin{aligned} \widetilde{Sponsorship}_{ijt} &= \beta_0 + \beta_1 \times 1\left(X_{it}^{followers} - c_m^{followers} > 0\right) \\ &+ \beta_2 \times 1\left(X_{it}^{avg.views} - c_m^{avg.views} > 0\right) + \vec{\Gamma} * \overrightarrow{ControlVariables} \\ &+ I_i + \tau_t + \epsilon_{ijt} \end{aligned} \tag{6}$$

$$Sentiment_{ijt} = \beta_0 + \beta_1 \times \widetilde{Sponsorship}_{ijt} + \beta_2 \times \overrightarrow{Voice_{ijt}^k} + \beta_3 \times \widetilde{Sponsorship}_{ijt} \times \overrightarrow{Voice_{ijt}^k} + f\left(X_{it}^{followers} - c_m^{followers}\right) + f\left(X_{it}^{avg.views} - c_m^{avg.views}\right) + \vec{\Gamma} * \overrightarrow{ControlVariables} + I_i + \tau_t + \epsilon_{ijt}, \epsilon_{ijt} \sim N(0, \sigma_\epsilon^2) \tag{7}$$

where the sponsorship indicator $\widetilde{Sponsorship}_{ijt}$ is instrumented by two binary indicators $\left(1\left(X_{it}^{followers} - c_m^{followers} > 0\right), 1\left(X_{it}^{avg.views} - c_m^{avg.views} > 0\right)\right)$ indicating influencer i's eligibility for sponsorship by brand m at time t. The assignment variables $X_{it}^{followers}$ and $X_{it}^{avg.views}$, representing follower count and average views, respectively, are subjected to thresholds $c_m^{followers}$ and $c_m^{avg.views}$, respectively. We employ the instrumented sponsorship variable in Equation (6) to estimate the effects of sponsorship, vocal characteristics, and their interaction on sentiment in Equation (7), using a cubic functional form. We verify the optimal polynomial order based on Akaike's criterion and F-statistics and conduct a sensitivity analysis

[7] As a robustness check, we also run the first-stage equation using the RD approach.

by dropping the outermost 1%, 5%, and 10% of the data. (See Web Appendix 4.D, 4.E, and 4.F for details.)

TWFE approach.

TWFE (Two-Way Fixed Effects) estimators enable the correlation of treatment with unobserved confounders, specific to influencers and time, concurrently (Xu 2017). This approach can identify the treatment exposure's time-varying evolution instead of a constant treatment effect (Chaisemartin and d'Haultfoeuille 2020). To specify the model, we use the approach outlined by Imai and Kim (2021):

$$Sentiment_{it} = \beta_0 + \beta_1 \times \widetilde{Sponsorship}_{it} + \beta_2 \times \overrightarrow{Voice^k_{it}} + \beta_3 \times \widetilde{Sponsorship}_{it} \times \overrightarrow{Voice^k_{it}} + \vec{\Gamma} \times \overrightarrow{ControlVariables} + I_i + \tau_t + \epsilon_{it}, \epsilon_{it} \sim N(0, \sigma^2_\epsilon)$$

$$\text{where } I_i = g(U_i), \tau_t = h(W_t) \tag{8}$$

$$\hat{\beta} = argmin_\beta \sum_{i=1}^{N} \sum_{t=1}^{T} [\{(Y_{it} - \bar{Y}) - (\bar{Y}_i - \bar{Y}) - (\bar{Y}_t - \bar{Y})\} - \beta\{(Z_{it} - \bar{Z}) - (\bar{Z}_i - \bar{Z}) - (\bar{Z}_t - \bar{Z})\}]^2$$

where U_i, W_t are unobserved confounders by influencer and time. The functional form g(.) and h(.) capture the influencer and time fixed effects. Influencer-specific means ($\bar{Y}_i = T^{-1} \sum_{t=1}^{T} Y_{it}, \bar{Z}_i = T^{-1} \sum_{t=1}^{T} Z_{it}$), time-specific means ($\bar{Y}_t = N^{-1} \sum_{i=1}^{N} Y_{it}, \bar{Z}_t = N^{-1} \sum_{i=1}^{N} Z_{it}$), and overall means ($\bar{Y} = N^{-1} \sum_{i=1}^{N} Y_{it}, \bar{Z} = N^{-1} \sum_{i=1}^{N} Z_{it}$) are calculated. The outcome variable (consumer sentiment) is denoted as Y, and sponsorship is represented by Z.

Comparison of the Three Identification Strategies

In our paper, it is important to acknowledge that the three identification strategies utilized in our analysis are based on distinct assumptions. Firstly, in the IV approach, we use two instruments that meet the exclusion restriction condition and demonstrate a significant correlation with sponsorship receipt but not with consumer sentiment. Secondly, in the FRD approach, we make the assumption of (1) monotonicity, meaning that influencers who pass the assignment variable cutoff do not have their probability of accepting or rejecting sponsorship impacted, and (2) excludability, meaning that the passing of the cutoff does not affect consumer sentiment except through sponsorship receipt (Lee and Lemieux, 2010). Finally, in the TWFE approach, we incorporate influencer and time fixed effects in order to estimate causal effects in our panel data by excluding the unknown influencer or time-specific heterogeneity.

The three Identification strategies utilize different data samples, which may account for differences in estimated treatment effects. Specifically, the IV approach employs a propensity score-matched sample that includes an equal number of sponsored and non-sponsored videos (50% each), while the TWFE approach utilizes all available videos (16% sponsored and 84% non-sponsored). Consequently, the IV results may overestimate the impact of sponsorship on consumer sentiment relative to the TWFE results. Furthermore, the FRD approach employs a sample of only two videos per influencer, one just before and one just after the assignment variable cutoffs. Therefore, the FRD results may only reflect the local average treatment effect around the cut-off point of the assignment variables, and its magnitude may be less generalizable than the IV and TWFE results.

Individual Heterogeneity in the Use of the Voice Modulation Strategy

Our third research question aims to identify the influencers who tend to adapt their vocal characteristics in sponsored videos to improve consumer sentiment. We exclusively examine influencer-specific characteristics, including age, influencer tenure (i.e., the duration for which they have been posting videos), and three popularity metrics (i.e., number of followers, total views, and Instagram followers). Specifically, we investigate whether influencers whose age, tenure, and popularity fall below the median are more likely to modulate their voice in an effort to offset any potential negative impact of sponsorship.

To investigate the impact of influencer characteristics on the average loudness in sponsored videos, we estimate a regression model that includes sponsorship, content characteristics, and interaction terms between sponsorship and influencer characteristics as independent variables. We focus on average loudness as the dependent variable as it is the only vocal characteristic significantly and robustly affecting consumer sentiment across the three identification strategies: IV, FRD and TWFE models. Equation (9) shows the formal model of estimating the relationship between sponsored influencer characteristics and the average loudness.

$$\begin{aligned} Voice^{*}_{ijt} = \ & \beta_0 + \overrightarrow{\beta_1} \times \overrightarrow{ContentControls_{ijt}} \\ & + \overrightarrow{\beta_2} \times Sponsorship_{ijt} \times \overrightarrow{ContentControls_{ijt}} + \epsilon_{ijt}, \epsilon_{ijt} \\ & \sim N(0, \sigma_{\epsilon}^{2}) \end{aligned} \tag{9}$$

RESULTS

Do Influencers Change Their Vocal Characteristics in Sponsored Videos?

Table 8 presents the estimation results for Equation (2), which demonstrates that influencers significantly modulate all five vocal characteristics in sponsored videos. Our findings reveal a significant decrease of 0.069 units in average loudness in sponsored videos compared to non-sponsored videos, which translates to a 4.47 dB decrease (6.9% of the average loudness, 64.69 dB) after accounting for standardization across individuals. Considering a conservative measure of the just-noticeable difference (JND) of vocal loudness at 1 dB, the observed change can be deemed noticeable (Ellingson, 2018).

We report significant effects of sponsorship on various vocal characteristics in sponsored videos. Specifically, influencers modulate their average pitch by 11.82 Hz (6.3% of the average pitch, 187.63 Hz) higher in sponsored videos, increase their loudness variability, decrease their pitch variability, and have longer talking duration in sponsored videos.

Table 8. Change in Vocal Characteristics in Sponsored Videos (Relative to Non-Sponsored Videos)

	(1)	(2)	(3)	(4)	(5)
Dependent Variable (DV)	Average Loudness	Average Pitch	Loudness Variability	Pitch Variability	Talking Duration
Sponsorship	-0.069***	0.063***	0.026**	-0.04***	0.097***
	(0.01)	(0.01)	(0.012)	(0.01)	(0.009)
Individual/time/content FEs	Yes	Yes	Yes	Yes	Yes
Observations (Individual)	32,920 (1,017)	32,920 (1,017)	32,920 (1,017)	32,920 (1,017)	32,920 (1,017)
R^2	0.346	0.321	0.195	0.304	0.425
Notes. Full model results can be found in Online Appendix 5.A.					

We present additional analysis of the effects of sponsorship on the other vocal characteristics (i.e., min and max of the vocal loudness) in Web Appendices 5.B and 5.C. Of the 985 vocal statistics analyzed, we find that 73.7% (726) are significant, further demonstrating the significant impact of sponsorship on influencers' vocal characteristics.

Do Vocal Changes in Sponsored Videos Affect Consumer Sentiment?

Tables 9 and 10 present the results of the IV, FRD, and TWFE models, which consistently indicate that the reduction of average loudness in sponsored videos significantly enhances consumer sentiment toward the sponsored content. In the IV model (Table 9), the negative main effect of sponsorship on consumer sentiment is statistically significant (-0.0059). However, a one standard deviation decrease in the average loudness of sponsored videos leads to an increase in consumer sentiment by 0.0026 (-0.0027 + 0.0053 = 0.0026). This finding suggests that lowering the average loudness in sponsored videos can counteract 49% (= 0.0026 / 0.0053) of the negative impact of sponsorship on consumer sentiment.

Table 9. Effects of Sponsorship, Vocal Characteristics, and Interactions on Sentiment: IV Model

	(1) IV Model	
Vocal characteristics	Coef.	SE
Average loudness	0.0027**	(0.0013)
Average pitch	0.0061	(0.0039)
Loudness variability	-0.0023	(0.0017)
Pitch variability	-0.0026	(0.0032)
Talking duration	-0.0088**	(0.0039)
Sponsorship & Sponsored Vocal Characteristics		
Sponsorship	-0.0059**	(0.0027)
*Sponsorship * Average loudness*	-0.0053**	(0.0025)
*Sponsorship * Average pitch*	-0.0072	(0.0068)
*Sponsorship * Loudness variability*	-0.0028	(0.0031)
*Sponsorship * Pitch variability*	0.0011	(0.0056)
*Sponsorship * Talking duration*	0.0046	(0.0069)
Constant	0.7606***	(0.0846)
Fixed Effects	Influencer, time, and content control characteristics	
Observation (Individual)	32,920 (1,017)	
R^2	0.3355	
Notes. Full model results can be found in Online Appendix 5.D. *p<0.1;**p<0.05;***p<0.01		

The results of the FRD model (Table 10, Model (1)) indicate that a reduction of one unit of average loudness in sponsored videos can mitigate the negative impact of sponsorship on consumer sentiment by 0.007 (-0.0118 + 0.0188 = 0.007). The TWFE model results (Table 10, Model (2)) further confirm that lowering average loudness in sponsored videos can mitigate the damage of sponsorship to consumer sentiment by 12% (=(-0.0018+0.0034) / |-0.0139|). It should be noted that the FRD approach only includes two videos per influencer, shortly before and after the assignment variable cutoffs per each brand sponsorship, whereas the TWFE approach includes all videos available. Therefore, the treatment effect estimated by FRD may only reflect the local average treatment effect around the cut-off point of the assignment variables.

Table 10. Effects of Sponsorship, Vocal Characteristics, and Interactions on Sentiment: FRD & TWFE Models

	(1) FRD Model		(2) TWFE Model	
	DV: Consumer Sentiment		DV: Consumer Sentiment	
Vocal characteristics	Coef.	SE	Coef.	SE
Average loudness	0.0118**	(0.0050)	0.0018***	(0.0006)
Average pitch	0.0155	(0.0158)	0.0023	(0.0018)
Loudness variability	-0.0107*	(0.0055)	-0.0016*	(0.0008)
Pitch variability	-0.0142	(0.0109)	0.0005	(0.0014)
Talking duration	-0.0328**	(0.0139)	-0.0056***	(0.0017)
Sponsorship & Sponsored Vocal Characteristics				
Sponsorship	0.0051	(0.0147)	-0.0139***	(0.0018)
*Sponsorship * Average loudness*	-0.0188**	(0.0086)	-0.0034**	(0.0016)
*Sponsorship * Average pitch*	-0.0230	(0.0181)	-0.0035	(0.0043)
*Sponsorship * Loudness variability*	0.0147	(0.0090)	-0.0030*	(0.0018)
*Sponsorship * Pitch variability*	0.0272*	(0.0152)	0.00001	(0.0034)
*Sponsorship * Talking duration*	0.0356**	(0.0180)	0.0027	(0.0042)
Constant	1.1847***	(0.3010)		
Fixed Effects	Influencer, time, and content control characteristics			
Observation (Individual)	2,514 (589)		103,479 (1,079)	
R^2	0.5637		0.3233	
Notes. Full model results can be found in Online Appendix 5.E.				

Our analysis indicates that while sponsorship significantly affects four vocal characteristics (Tables 9 & 10), only average loudness has a consistent impact on consumer sentiment across all three models. Specifically, sponsorship adversely affects consumer sentiment, but reducing average loudness in sponsored videos mitigates the negative impact of sponsorship. This finding underscores the importance of the appropriate level of loudness in business contexts, which has been debated in the literature. Although one study reported that higher loudness reduces refusal rates in telephone surveys (Oksenberg et al. 1986), others found that moderate loudness is more credible than high loudness in banking telemarketing (Chebat et al. 2007). Our results support the latter finding, suggesting that moderate loudness is more effective for persuasion in influencer marketing.

What Kinds of Influencers Decrease Their Average Loudness in Sponsored Videos?

Table 11 presents the results of the estimation of Equation (9), which identifies the key influencer characteristics that predict a significant change in the average loudness of sponsored videos (compared to non-sponsored videos). The reported results display the interaction coefficients between sponsorship and the influencer subsets constructed via influencer-specific characteristics: age, tenure and three popularity metrics.

Table 11. Predictors of a Significant Change in Average Loudness in Sponsored Videos

	DV: Average Loudness	
Variable	Coef.	SE
Sponsorship	-0.059***	(0.018)
Interaction Variables between Sponsorship & Content Control Variables		
*Sponsorship*1(age ≤ median age)*	-0.001	(0.017)
*Sponsorship*1(tenure ≤ median tenure)*	-0.003	(0.018)
*Sponsorship*1(n. of followers ≤ median n. of followers)*	-0.049**	(0.020)
*Sponsorship*1(n. of total views ≤ median n. of total views)*	-0.0004	(0.018)
*Sponsorship*1(n. of Insta. followers ≤ median n. of Insta. followers)*	0.032	(0.020)
Constant	1.538*	(0.880)
Observations (Individual)	32,920 (1,017)	
R^2	0.3462	
Notes. Full model results can be found in Online Appendix 5.F. *p<0.1;**p<0.05;***p<0.01		

Our results show that influencers with a smaller following tend to decrease their average vocal loudness in sponsored videos. In our analyses, we considered several influencer-specific characteristics, including age, tenure, and the number of total views and Instagram followers. However, only the subset of influencers categorized by the number of followers demonstrated a significant association with the sponsorship indicator.

The current analysis involves numerous covariates, which could potentially produce spurious results. Given the scale of the analysis, we cannot entirely rule out the possibility of five false-positive results out of 100. The unreported coefficients are reported in Web Appendix 5.F.

DISCUSSION, ROBUSTNESS AND VALIDITY CHECK

In the supplementary material, Web Appendix 6.A reports the IV model results with additional instruments for vocal characteristics and Web Appendix 6.B shows the IV model results with interactions between sponsorship and vocal and visual content variables. Web Appendix 6.C presents an analysis using a different dependent variable, the number of future sponsorship deals, and reveals a significant relationship between vocal characteristics and the influencers' future sponsorship deals from the same brand, which supports our main finding about the impact of vocal characteristics on consumer sentiment. Additionally, in Web Appendix 6.D, we provide evidence that influencers who increase their loudness in sponsored videos are more negatively affected by sponsorship disclosure.

We present robustness and validity checks in the Web Appendices to reinforce the findings of our research. For validity checks, we assess the possibility of reverse causality (Web Appendix 6.E), which would suggest that brands give sponsorship to influencers who they know would strategically reduce loudness in response. However, our results indicate that this alternative interpretation of reverse causality is not supported. We also perform validity checks for the FRD model to show that assignment variables were not manipulated (Web Appendix 6.F) and use placebo thresholds (Web Appendix 6.G). For robustness checks, we report consistent effects of sponsorship and voice-sponsorship interactions (1) with unmatched data with progressively added control variables (Web Appendix 6.H), (2) with the addition of more advertising media control variables (Web Appendix 6.I).

CONCLUSION, IMPLICATIONS, AND FUTURE RESEARCH

This paper investigates the impact of sponsorship on influencer vocal characteristics in videos and its effect on consumer sentiment. Our study reveals that influencers lower their average loudness in sponsored videos, which partially mitigates the negative impact of sponsorship on consumer sentiment. Our results are consistent with prior research indicating that moderate loudness appears more credible than high loudness (Chebat et al. 2007). We find that influencers with a smaller following tend to decrease their average vocal loudness in sponsored videos.

Multiple accounts could explain the differences in vocal characteristics between sponsored and non-sponsored videos. One possible explanation is that optimal voice attribute levels differ between the two video types, with a louder voice being more effective in non-sponsored videos, while a quieter voice is preferred in sponsored videos. Additionally, influencers may receive coaching or training from brands for sponsored videos, resulting in a more artificial vocal style, while they may use a more natural style in non-sponsored videos. Finally, influencers may choose to modulate their voice to regain consumer trust in sponsored videos, as voice modulation may require effort, while non-sponsored videos lack the same degree of scrutiny.

Our research has limitations, and evidence of influencers' awareness of their vocal adjustments is beyond the scope of this paper, leaving room for future research. However, our findings have implications for influencers, firms, consumers, and regulatory bodies. Influencers may benefit from intentionally speaking softer in sponsored videos to mitigate the negative impact of sponsorship on consumer sentiment. Firms should consider this strategy to enhance the success of their influencer marketing campaigns. Consumers should be aware that influencers may adjust their voice as a strategy to generate positive sentiment toward sponsored videos, and regulatory bodies may need to develop policies to better inform consumers about this tactic.

GENERAL ACKNOWLEDGEMENTS

We acknowledge the valuable contribution of the Facebook CrowdTangle team in providing us with the Instagram statistics data for this research.

REFERENCES

Barker S (2018), "A detailed strategy to find the right influencers for your business," (Accessed October 1, 2019), https://medium.com/swlh/a-detailed-strategy-to-find-the-right-influencers-for-your-business-9e3dc4e511a1.

Banai, I. P., Banai, B., & Bovan, K. (2017). Vocal characteristics of presidential candidates can predict the outcome of actual elections. *Evolution and Human Behavior*, 38(3), 309-314.

Batra, R. (2019). Creating brand meaning: A review and research agenda. *Journal of Consumer Psychology*, 29(3), 535-546.

Biron, B (2019), “Beauty has blown up to be a $532 billion industry — and analysts say that these 4 trends will make it even bigger,” Business Insider (Accessed November 6, 2020)

Brysbaert, Marc, Amy B. Warriner, and Victor Kuperman (2014), “Concreteness ratings for 40,000 generally known English word lemmas,” *Behavior research methods*, 46 (3), 904-911.

Chakraborty, Rupayan, et al. (2015), “TCS-ILAB-MediaEval 2015: Affective Impact of Movies and Violent Scene Detection,” *MediaEval*.

Chebat, Jean-Charles, et al. (2007), “Voice and persuasion in a banking telemarketing context,” *Perceptual and motor skills*, 104 (2), 419-437.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964-96.

Eyben, Florian, Martin Wöllmer, and Björn Schuller (2010), “Opensmile: the munich versatile and fast open-source audio feature extractor.” *Proceedings of the 18th ACM international conference on Multimedia*.

Fast Company (2019), “Sephora picks 24 influencers for its coveted #SephoraSquad program,” (Accessed September 17, 2019).

Fong, H., Kumar, V., & Sudhir, K. (2021). A theory-based interpretable deep learning architecture for music emotion. Available at SSRN 4025386.

Fraccaro, Paul J., et al. (2011), “Experimental evidence that women speak in a higher voice pitch to men they find attractive,” *Journal of Evolutionary Psychology* 9 (1), 57-67.

FTC (2019), “The FTC's endorsement guides: Being up-front with consumers,” (Accessed September 10, 2019).

FTC (2020), "FTC bureau consumer protection," (Accessed October 20, 2020), https://www.ftc.gov/about-ftc/bureaus-offices/bureau-consumer-protection

Gents Scents (2020), "The TRUTH About ScentBird," YouTube, (Accessed December 26, 2018), https://www.youtube.com/watch?v=GPPR25boTW4

Globe Newswire (2021), "Collabstr Releases 2022 Influencer Marketing Report With Key Sector Data" (Accessed March 08 2022), https://www.globenewswire.com/news-release/2021/12/07/2347692/0/en/Collabstr-Releases-2022-Influencer-Marketing-Report-With-Key-Sector-Data.html.

Grootendorst, M. (2020). Bertopic: Leveraging BERT and C-Tf-Idf to create easily interpretable topics. Version v0, 4.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

Hutto, Clayton, and Eric Gilbert, (2014) "Vader: A parsimonious rule-based model for sentiment analysis of social media text." *Proceedings of the International AAAI Conference on Web and Social Media*, 8 (1).

Imai, K., & Kim, I. S. (2021). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, 29(3), 405-415.

Kirmani, A., & Zhu, R. (2007). Vigilant against manipulation: The effect of regulatory focus on the use of persuasion knowledge. Journal of Marketing Research, 44(4), 688-701.

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. Proceedings of the Royal Society B: Biological Sciences, 279(1738), 2698-2704.

Kozinets, Robert V., et al. (2010), “Networked narratives: Understanding word-of-mouth marketing in online communities,” *Journal of marketing*, 74 (2), 71-89.

Krause, Amanda (2020) “Inside the YouTube beauty community that's turning makeup artists into millionaires,” (Accessed November 27, 2020).

Lee, David S., and Thomas Lemieux (2010), “Regression discontinuity designs in economics,” *Journal of economic literature*, 48 (2), 281-355.

Lawson, C. E. (2021). Skin deep: Callout strategies, influencers, and racism in the online beauty community. *New Media & Society*, 23(3), 596-612.

Li, T., Qian, R., Dong, C., Liu, S., Yan, Q., Zhu, W., & Lin, L. (2018, October). Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In Proceedings of the 26th ACM international conference on Multimedia (pp. 645-653).

Liang, L., Lin, L., Jin, L., Xie, D., & Li, M. (2018). SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. In 2018 24th International Conference on Pattern Recognition (ICPR) (pp. 1598-1603). IEEE.

Lou, C., & Yuan, S. (2019). Influencer marketing: how message value and credibility affect consumer trust of branded content on social media. Journal of Interactive Advertising, 19(1), 58-73.

Mehrabian, A. (2017). Nonverbal communication. Routledge.

Mobius, M. M., & Rosenblat, T. S. (2006). Why beauty matters. American Economic Review, 96(1), 222-235.

Octoly (2015), “Who are the Top 5 beauty brands on YouTube?,” (Accessed February 13, 2019).

Oksenberg, L., Coleman, L., & Cannell, C. F. (1986). Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly*, 50(1), 97-111.

Papay, John, John Willett, and Richard Murnane (2011), “Extending the regression-discontinuity approach to multiple assignment variables,” *Journal of Econometrics*, 161 (2), 203-207.

Peterson, Robert A., Michael P. Cannito, and Steven P. Brown (1995), “An exploratory investigation of voice characteristics and selling effectiveness,” *Journal of Personal Selling & Sales Management,* 15 (1), 1-15.

Scherer, Klaus, Harvey London, and Jared J. Wolf (1973), “The voice of confidence: Paralinguistic cues and audience evaluation,” *Journal of Research in Personality*, 7 (1), 31-44.

Schneider, Matthew J., and Sachin Gupta (2016) “Forecasting sales of new and existing products using consumer reviews,” *International Journal of Forecasting*, 32 (2), 243-256.

Schuller, Björn, et al. (2019) “Affective and behavioural computing: Lessons learnt from the first computational paralinguistics challenge,” *Computer Speech & Language* 53, 156-180.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, 27(5), 499-520.

Schwemmer, Carsten, and Sandra Ziewiecki (2018), “Social media sellout: The increasing role of product promotion on YouTube,” *Social Media+ Society*, 4 (3).

Statista Search Department (2021, August 13th) Share of influencers posting sponsored content on Instagram 2019, by gender [Infographic]. Statista. https://www.statista.com/statistics/893749/share-influencers-creating-sponsored-posts-by-gender/

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708).

Talebi, H., & Milanfar, P. (2018). NIMA: Neural image assessment. IEEE transactions on image processing, 27(8), 3998-4011.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1), 57-76.

Wiener, Hillary JD, and Tanya L. Chartrand (2014), "The effect of voice quality on ad efficacy," *Psychology & Marketing*, 31 (7), 509-517.

WWD (2016), "Women Daily," (Accessed February 13, 2019), https://wwd.com/beauty-industry-news/beauty-features/loreal-loreal-league-bloggers-forms-loreal-league-10473153.

Yang, J., Xie, Y., Krishnamurthi, L., & Papatla, P. (2022). High-energy ad content: A large-scale investigation of tv commercials. Journal of Marketing Research, 59(4), 840-859.

Zuckerman, Miron, and Robert E. Driver (1989), "What sounds beautiful is good: The vocal attractiveness stereotype," *Journal of nonverbal behavior*, 13 (2), 67-82.

VOICE ANALYTICS OF ONLINE INFLUENCERS

ONLINE APPENDIX

Table of Contents

OA SECTION 1. FOUR CASE ANALYSES TO ADDRESS THE POTENTIAL NON-DISCLOSURE PROBLEM

OA 1.A. SPONSORSHIP DISCLOSURE

We acknowledge that in our paper, we consider all videos with a sponsorship disclosure as sponsored videos and all videos without such disclosure as non-sponsored. However, it is plausible that some non-disclosed videos are actually sponsored and lack proper disclosure. These videos may include brand names or affiliation codes without any sponsorship disclosure. To address this, we designate videos with possible indications of sponsorship but lacking disclosure as "uncertain videos."

Table S1 categorizes the videos into three categories: disclosed ("D"), uncertain ("U"), and non-sponsored ("N"). Disclosed videos (D) are those where the influencer explicitly discloses sponsorship in the video description box. Uncertain videos (U) are those where the influencer does not disclose sponsorship but may have mentioned a sponsoring brand or included an affiliation code in the description. Non-sponsored videos (N) are those where the influencer does not disclose sponsorship and does not include brand or discount information in the description.

Table S2 compares four cases to examine the impact of uncertain videos on the paper's conclusions. Case 1 considers complete honesty, in which uncertain videos are treated as non-sponsored. Thus, the treatment group consists of disclosed videos, while the control group comprises uncertain and non-sponsored videos. Cases 2 and 3 assume complete dishonesty, where all uncertain videos are sponsored but not disclosed properly. In Case 2, disclosed and uncertain videos comprise the treatment group, and non-sponsored videos are the control group. In Case 3, only uncertain videos serve as the treatment group, and non-sponsored videos serve as the control group. Case 4 eliminates uncertainty by using only disclosed videos as the treatment group and non-sponsored videos as the control group.

In Case 1, we found that sponsorship significantly reduces consumer sentiment compared to the control group, whereas an increase in average loudness significantly improves it, and a decrease in average loudness in sponsored videos partially mitigates the negative effect of sponsorship on consumer sentiment (refer to Tables 9 and 10 in the main paper). These findings are consistent with those of Case 2 (Table S3), indicating that our primary conclusions remain valid even if we misclassified all uncertain videos.

Case 3 results, as presented in Table S4, show an insignificant effect of sponsorship on consumer sentiment and an insignificant interaction effect of sponsorship and average loudness. These findings align with the expectations of the study since if uncertain videos are indeed non-sponsored, there would be no need for influencers to adjust their voices to mitigate any negative effect of sponsorship. Alternatively, if uncertain videos are sponsored but undisclosed, consumers would not be affected by the disguised sponsorship and hence, influencers would not need to adjust their voices. These insignificant results do not challenge the main conclusions of the study.

The results of Case 4, presented in Table S5, support the main conclusion of the paper. Thus, the case studies suggest that the paper's findings are unaffected by the potential issue of non-disclosure.

Table S1. Video Classification Based on Disclosure and Sponsorship Status

Disclosure Status	**Disclosed**	**Non-disclosed**	
Sponsorship Status	**Sponsored**	**Uncertain**	**Non-sponsored**
Category Label	**Disclosed (D)**	**Uncertain (U)**	**Non-sponsored (N)**
Description Content	Explicit disclosure	No disclosure; brand mention and/or affiliate code	No disclosure; Neither brand mention nor affiliate code
Description Example	"This video is sponsored by L'Oreal Paris."	"I am going to use L'Oreal lip gloss…." "Use code MKT and get 15% off"	"This video is not sponsored."
Notes. The video classification in the study employs L'Oreal Paris as a brand sponsor case. The uncertain video category (U) includes examples of brand mentions and affiliate codes. The affiliate code example assumes that "MKT" is a distinct code for the influencer's discount affiliate codes.			

Table S2. Four Cases to Examine Robustness Against Non-Disclosed Sponsorship

Case	**Treatment**	**Control**	**Difference in Sentiment**
Case 1	D	U+N	Significant
Case 2	D+U	N	Significant
Case 3	U	N	Not significant
Case 4	D	N	Significant

OA 1.B. RESULTS OF CASE 2

Table S3 displays the findings from Case 2, which assumes complete dishonesty, where disclosed and uncertain videos represent the treatment group, and non-sponsored videos act as the control group. The findings are in line with the main results, as average loudness has a positive, significant coefficient (0.0111), sponsorship has a negative, significant coefficient (-0.0125), and the interaction between average loudness and sponsorship has a negative, significant coefficient on consumer sentiment (-0.0136). These results demonstrate that the primary results are valid even if all uncertain videos are misclassified.

Table S3. Results of Case 2: Treatment (Disclosed + Uncertain) vs. Control (Non-sponsored)

	DV: Consumer Sentiment

Variable	Coef.	SE
Average loudness	**0.0111****	**(0.0047)**
Average pitch	**0.0140**	**(0.0105)**
Loudness variability	**0.0032**	**(0.0061)**
Pitch variability	**-0.0039**	**(0.0085)**
Talking duration	**-0.0137**	**(0.0103)**
Sponsorship & Sponsored Vocal Characteristics		
Sponsorship	**-0.0125****	**(0.0052)**
Sponsorship * Average loudness	**-0.0136****	**(0.0061)**
Sponsorship * Average pitch	**-0.0151**	**(0.0140)**
Sponsorship * Loudness variability	**-0.0081**	**(0.0077)**
Sponsorship * Pitch variability	**0.0023**	**(0.0114)**
Sponsorship * Talking duration	**0.0097**	**(0.0138)**
Other Content Control Variables		
Duration	0.0030*	(0.0018)
View count	-0.0245***	(0.0022)
Like count	0.0331***	(0.0030)
Dislike count	-0.0172***	(0.0015)
Comment count	-0.0078***	(0.0018)
Title length	-0.0042*	(0.0024)
Description length	0.0086***	(0.0019)
Tenure	0.0017	(0.0044)
Raw followers	-0.0039	(0.0031)
Raw total views	-0.0007	(0.0035)
Raw Instagram followers	-0.0020	(0.0020)
Avg. diff. in followers	-0.0012	(0.0008)
Avg. diff. in total views	-0.00001	(0.0011)
Avg. diff in Instagram followers	0.0003	(0.0005)
Format - GRWM	0.0065***	(0.0020)
Format - haul	0.0061***	(0.0019)
Format - review	-0.0022	(0.0020)
Format - routine	-0.0054***	(0.0019)
Format - tutorial	-0.0020	(0.0020)
Format - vlog	0.0012	(0.0020)
Verbal sophistication 1gram	-0.0077***	(0.0029)
Verbal sophistication 2gram	0.0065	(0.0050)
Verbal sophistication 3gram	0.0037	(0.0038)
Verbal concreteness	0.0104	(0.0185)
Verbal topic 0	-0.0002	(0.0032)
Verbal topic 1	0.0029	(0.0055)
Verbal topic 2	0.0095	(0.0078)
Verbal topic 3	0.0166***	(0.0045)
Verbal topic 4	-0.0178***	(0.0062)
Verbal topic 5	-0.0147**	(0.0065)
Verbal topic 6	0.0035	(0.0074)

Verbal topic 7	-0.0192**	(0.0087)
Verbal topic 8	0.0189**	(0.0080)
Verbal topic 9	-0.0199**	(0.0098)
Age	0.0032	(0.0074)
Emotion - angry	-0.0058	(0.0048)
Emotion - disgust	0.0117	(0.0088)
Emotion - fear	0.0006	(0.0039)
Emotion - happy	0.0069***	(0.0024)
Emotion - sad	0.0023	(0.0032)
Emotion - surprise	-0.0022	(0.0044)
Race - asian	0.0140**	(0.0058)
Race - black	0.0086	(0.0067)
Race - indian	0.0184*	(0.0099)
Race - latino_hispanic	0.0119	(0.0089)
Race - middle_eastern	0.0036	(0.0075)
Facial attractiveness	-0.0019	(0.0092)
Makeup heaviness	-0.0172	(0.0292)
Object – top/t-shirt/sweatshirt	0.0057	(0.0052)
Object – jacket		
Object – pants	0.0070	(0.0102)
Object – skirt	-0.0031	(0.0538)
Object – dress	0.0012	(0.0019)
Object - glasses	-0.0163	(0.0109)
Object – hat	0.0498	(0.0398)
Object – watch	-0.0318*	(0.0192)
Object – belt	0.0291	(0.0441)
Object – tights/stockings	-0.0205	(0.0212)
Object – shoe	0.0050**	(0.0025)
Object – bag/wallet	-0.0366	(0.0381)
Object – collar	0.0078	(0.0214)
Object – lapel	0.0272	(0.0282)
Object – sleeve	-0.0017	(0.0016)
Object – pocket	0.0423	(0.0259)
Object – neckline	0.0001	(0.0017)
Foundation1_G	-0.0153*	(0.0090)
Foundation2_G	0.0063	(0.0110)
Foundation3_G	0.0131	(0.0090)
Blush1_G	0.0107	(0.0146)
Blush2_G	0.0046	(0.0223)
Blush3_G	-0.0397	(0.0305)
Blush4_G	0.0580	(0.0382)
Blush5_G	-0.0270	(0.0260)
Lip_G	-0.0036	(0.0073)
Lipliner_G	0.0004	(0.0109)
Eyeshadow1_G	-0.0061	(0.0079)

Eyeshadow2_G	0.0216	(0.0168)
Eyeshadow3_G	-0.0014	(0.0245)
Eyeshadow4_G	-0.0526*	(0.0304)
Eyeshadow5_G	0.0335	(0.0216)
Eyeliner_G	-0.0130	(0.0089)
Eyebrow_G	0.0031	(0.0075)
NIMA score average	-0.0023	(0.0028)
Monthly ad. spending	0.0001	(0.0003)
Constant	0.7796***	(0.0850)
Observations (individual)	32,920 (1,017)	
Fixed Effects	Individual, year-month (time), 30 brands, 30 product types	
R^2	0.3355	
Note: *p < 0.1; **p < 0.05; ***p < 0.01		

OA 1.C. RESULTS OF CASE 3

Table S4 presents the findings of Case 3, which evaluates 7,605 uncertain videos against 8,855 non-sponsored videos from a pool of 16,460 matched control videos. In this scenario, we modify the definition of the instrumental variable since none of the videos have any disclosed sponsoring brands. Therefore, we utilize the brands shown in the uncertain videos (categorized as treated group in Case 3) to measure the number of influencers that have featured the same brand in the same week (and could have been concurrently sponsored by the brand without disclosure).

Our study's main effect of sponsorship on consumer sentiment is insignificant, unlike in the primary analysis. Additionally, the interaction effect of sponsorship and average loudness is insignificant. Thus, it can be concluded that uncertain videos do not affect the main conclusion of the primary analysis that influencers reduce average loudness in sponsored videos to counteract the negative impact of sponsorship on consumer sentiment.

Table S4. Results of Case 3: Treatment (Uncertain) vs. Control (Non-sponsored)

	DV: Consumer Sentiment	
Variable	Coef.	SE
Average loudness	**-0.0318**	**(0.0358)**
Average pitch	**-0.1139**	**(0.1393)**
Loudness variability	**0.0100**	**(0.0179)**
Pitch variability	**0.0820**	**(0.1114)**
Talking duration	**0.0894**	**(0.1059)**
Sponsorship & Sponsored Vocal Characteristics		
Sponsorship	**-0.9838**	**(1.7026)**
Sponsorship * Average loudness	**0.0567**	**(0.0583)**

Sponsorship * Average pitch	**0.2256**	**(0.2545)**
Sponsorship * Loudness variability	**-0.0268**	**(0.0313)**
Sponsorship * Pitch variability	**-0.1444**	**(0.1791)**
Sponsorship * Talking duration	**-0.1900**	**(0.1970)**
Other Content Control Variables		
Duration	-0.0014	(0.0104)
View count	-0.0183	(0.0259)
Like count	0.0068	(0.0639)
Dislike count	-0.0179***	(0.0022)
Comment count	-0.0024	(0.0077)
Title length	-0.0112	(0.0101)
Description length	-0.0658	(0.1292)
Tenure	-0.0186	(0.0383)
Raw followers	-0.0094	(0.0073)
Raw total views	-0.0138	(0.0104)
Raw Instagram followers	0.0138	(0.0264)
Avg. diff. in followers	-0.0055	(0.0089)
Avg. diff. in total views	0.0019	(0.0040)
Avg. diff in Instagram followers	-0.0011	(0.0031)
Format - GRWM	0.0211	(0.0206)
Format - haul	0.0053	(0.0036)
Format - review	-0.0504	(0.0879)
Format - routine	-0.0191	(0.0240)
Format - tutorial	-0.0111	(0.0187)
Format - vlog	0.0060	(0.0085)
Verbal sophistication 1gram	-0.0183	(0.0266)
Verbal sophistication 2gram	0.0294	(0.0253)
Verbal sophistication 3gram	-0.0166	(0.0167)
Verbal concreteness	0.0207	(0.0840)
Verbal topic 0	0.0047	(0.0051)
Verbal topic 1	-0.0106	(0.0378)
Verbal topic 2	0.0460	(0.0655)
Verbal topic 3	0.0169	(0.0105)
Verbal topic 4	-0.0696	(0.0581)
Verbal topic 5	-0.0166	(0.0168)
Verbal topic 6	0.0039	(0.0090)
Verbal topic 7	0.0142	(0.0438)
Verbal topic 8	0.0313	(0.0284)
Verbal topic 9	-0.0528	(0.0583)
Age	-0.0294	(0.0441)

Emotion - angry	-0.0283	(0.0276)
Emotion - disgust	0.0381	(0.0659)
Emotion - fear	0.0046	(0.0077)
Emotion - happy	0.0102	(0.0085)
Emotion - sad	-0.0054	(0.0178)
Emotion - surprise	-0.0154	(0.0228)
Race - asian	0.0263	(0.0361)
Race - black	-0.0012	(0.0280)
Race - indian	0.0150	(0.0330)
Race - latino_hispanic	0.0389	(0.0429)
Race - middle_eastern	-0.0011	(0.0123)
Facial attractiveness	-0.0860	(0.1739)
Makeup heaviness	-0.1483	(0.1957)
Object – top/t-shirt/sweatshirt	-0.0182	(0.0406)
Object – jacket		
Object – pants	-0.0138	(0.0342)
Object – skirt	-0.0169	(0.0739)
Object – dress	-0.0023	(0.0056)
Object - glasses	0.0014	(0.0407)
Object – hat	0.0402	(0.1455)
Object – watch	-0.0409	(0.0431)
Object – belt		
Object – tights/stockings	-0.0343	(0.0460)
Object – shoe	0.0070*	(0.0040)
Object – bag/wallet	0.1186	(0.3014)
Object – collar	-0.0093	(0.0537)
Object – lapel	-0.0240	(0.0905)
Object – sleeve	0.0064	(0.0106)
Object – pocket	0.0500	(0.0464)
Object – neckline	0.0061	(0.0118)
Foundation1_G	0.0486	(0.0832)
Foundation2_G	-0.0773	(0.1281)
Foundation3_G	0.0436	(0.0487)
Blush1_G	-0.0808	(0.1409)
Blush2_G	0.1968	(0.2763)
Blush3_G	-0.2819	(0.2982)
Blush4_G	0.2794	(0.3273)
Blush5_G	-0.0656	(0.0937)
Lip_G	-0.0052	(0.0180)
Lipliner_G	0.0044	(0.0167)

Eyeshadow1_G	-0.0111	(0.0254)
Eyeshadow2_G	-0.1075	(0.2432)
Eyeshadow3_G	0.1345	(0.3188)
Eyeshadow4_G	-0.2578	(0.3784)
Eyeshadow5_G	0.1588	(0.1896)
Eyeliner_G	0.0101	(0.0396)
Eyebrow_G	0.0113	(0.0287)
NIMA score average	-0.0062	(0.0042)
Monthly ad. spending	-0.0030	(0.0054)
Constant	2.8659	(3.5252)
Observations (individual)	32,920 (1,017)	
Fixed Effects	Individual, year-month (time), 30 brands, 30 product types	
R^2	0.3825	
Note: *p < 0.1; **p < 0.05; ***p < 0.01		

OA 1.D. RESULTS OF CASE 4

Table S5 presents the findings of Case 4, which excludes uncertain videos from the analysis. The results confirm the main findings with a positive, significant coefficient for average loudness (0.0047), a negative, significant coefficient for sponsorship (-0.0103), and a negative, significant coefficient for the interaction between average loudness and sponsorship on consumer sentiment (-0.0075). These results indicate that the inclusion of uncertain videos did not significantly affect the conclusions drawn in the main paper.

Table S5. Results of Case 4: Treatment (Disclosed) vs. Control (Non-sponsored)

	DV: Consumer Sentiment	
Variable	Coef.	SE
Average loudness	**0.0047****	**(0.0022)**
Average pitch	**0.0032**	**(0.0064)**
Loudness variability	**-0.0019**	**(0.0031)**
Pitch variability	**-0.0003**	**(0.0053)**
Talking duration	**-0.0049**	**(0.0062)**
Sponsorship & Sponsored Vocal Characteristics		
Sponsorship	**-0.0103*****	**(0.0040)**
Sponsorship * Average loudness	**-0.0075****	**(0.0032)**
Sponsorship * Average pitch	**-0.0045**	**(0.0094)**
Sponsorship * Loudness variability	**-0.0029**	**(0.0043)**
Sponsorship * Pitch variability	**-0.0009**	**(0.0077)**
Sponsorship * Talking duration	**0.0014**	**(0.0091)**

Other Content Control Variables		
Duration	0.0023	(0.0020)
View count	-0.0225***	(0.0025)
Like count	0.0302***	(0.0035)
Dislike count	-0.0155***	(0.0017)
Comment count	-0.0088***	(0.0021)
Title length	0.0009	(0.0026)
Description length	0.0128***	(0.0025)
Tenure	0.0025	(0.0061)
Raw followers	-0.0031	(0.0037)
Raw total views	0.0025	(0.0037)
Raw Instagram followers	-0.0029	(0.0027)
Avg. diff. in followers	-0.0001	(0.0009)
Avg. diff. in total views	-0.0015	(0.0013)
Avg. diff in Instagram followers	0.0002	(0.0006)
Format - GRWM	0.0059***	(0.0022)
Format - haul	0.0044**	(0.0021)
Format - review	-0.0020	(0.0022)
Format - routine	-0.0057***	(0.0021)
Format - tutorial	-0.00002	(0.0022)
Format - vlog	0.0019	(0.0023)
Verbal sophistication 1gram	-0.0093***	(0.0033)
Verbal sophistication 2gram	0.0037	(0.0061)
Verbal sophistication 3gram	0.0106**	(0.0043)
Verbal concreteness	0.0117	(0.0214)
Verbal topic 0	0.0014	(0.0036)
Verbal topic 1	0.0010	(0.0062)
Verbal topic 2	0.0099	(0.0079)
Verbal topic 3	0.0159***	(0.0052)
Verbal topic 4	-0.0133**	(0.0063)
Verbal topic 5	-0.0132*	(0.0074)
Verbal topic 6	0.0020	(0.0084)
Verbal topic 7	-0.0226**	(0.0100)
Verbal topic 8	0.0222**	(0.0093)
Verbal topic 9	-0.0196*	(0.0110)
Age	0.0067	(0.0083)
Emotion - angry	0.0009	(0.0053)
Emotion - disgust	0.0108	(0.0102)
Emotion - fear	-0.0006	(0.0043)
Emotion - happy	0.0081***	(0.0027)

Emotion - sad	0.0025	(0.0036)
Emotion - surprise	0.0015	(0.0050)
Race - asian	0.0156**	(0.0066)
Race - black	0.0081	(0.0077)
Race - indian	0.0165	(0.0118)
Race - latino_hispanic	0.0181*	(0.0101)
Race - middle_eastern	0.0010	(0.0081)
Facial attractiveness	-0.0092	(0.0105)
Makeup heaviness	-0.0113	(0.0341)
Object – top/t-shirt/sweatshirt	0.0045	(0.0059)
Object – jacket		
Object – pants	0.0063	(0.0123)
Object – skirt	0.0534*	(0.0309)
Object – dress	0.0003	(0.0021)
Object - glasses	-0.0200*	(0.0115)
Object – hat	0.0611	(0.0440)
Object – watch	-0.0501***	(0.0185)
Object – belt	0.0184	(0.0338)
Object – tights/stockings	-0.0284	(0.0237)
Object – shoe	0.0037	(0.0028)
Object – bag/wallet	-0.0205	(0.0331)
Object – collar	0.0076	(0.0261)
Object – lapel	0.0294	(0.0331)
Object – sleeve	-0.0029	(0.0018)
Object – pocket	0.0314	(0.0307)
Object – neckline	0.0020	(0.0019)
Foundation1_G	-0.0142	(0.0103)
Foundation2_G	0.0136	(0.0122)
Foundation3_G	0.0081	(0.0104)
Blush1_G	0.0138	(0.0164)
Blush2_G	0.0017	(0.0238)
Blush3_G	-0.0028	(0.0338)
Blush4_G	0.0264	(0.0457)
Blush5_G	-0.0339	(0.0315)
Lip_G	-0.0031	(0.0081)
Lipliner_G	0.0054	(0.0122)
Eyeshadow1_G	-0.0098	(0.0090)
Eyeshadow2_G	0.0058	(0.0191)
Eyeshadow3_G	0.0303	(0.0279)
Eyeshadow4_G	-0.0712**	(0.0350)

Eyeshadow5_G	0.0357	(0.0249)
Eyeliner_G	-0.0217**	(0.0104)
Eyebrow_G	0.0022	(0.0083)
NIMA score average	-0.0011	(0.0030)
Monthly ad. spending	0.0003	(0.0003)
Constant	0.7793***	(0.1017)
Observations (individual)	32,920 (1,017)	
Fixed Effects	Individual, year-month (time), 30 brands, 30 product types	
R^2	0.3444	

OA SECTION 2. SUMMARY STATISTICS OF VOICES & CONTROL VARIABLES

OA 2.A. FIVE VOCAL CHARACTERISTICS

We utilized OpenSMILE, a cutting-edge speech recognition package (Eyben et al. 2010; Eyben, 2015), to analyze the sound files. OpenSMILE segments the files into brief time frames (in our study, 0.0125 seconds) and calculates vocal feature values within each frame, from which it derives various statistics such as the mean and standard deviation of the vocal features (including average loudness, average pitch, loudness variability, pitch variability, and talking duration). We employed the "emobase" configuration file to extract prosodic features (loudness, pitch, loudness variability, and pitch variability) as well as a voice quality feature (talking duration). To describe the five vocal features, we rely heavily on Eyben (2015) and Eyben et al. (2010).

Average loudness (intensity)

The amount of energy in a sound, known as average loudness or intensity, can be measured by a narrow band approximation of the voice signal (Kießling 1997; Eyben 2015). OpenSMILE measures average loudness in watts per square meter, which can be converted to decibels (dB), the more commonly used unit of loudness, as follows:

$$Vocal\ intensity\ (dB) = 10 \times \log_{10} \frac{Vocal\ intensity\ (Watts/meter^2)}{I_0} \tag{S1}$$
$$where\ the\ reference\ intensity\ I_0 = 1 \times 10^{-12} W/m^2$$

where watts is a unit of power, and meter^2 is a unit of area. The unit of power, watts, and the unit of area, meter^2, together indicate the sound power in a given unit of area as watts per square meter.

Average pitch (F0)

Pitch is a fundamental acoustic characteristic of speech that is measured in Hz. When a person's voice has a relatively high fundamental frequency, the speaker is perceived as having a higher pitch. The average speaking pitch for an adult male is 100 to 120 Hz, while for an adult female, it is 200 to 220 Hz (Simpson 2009).

Loudness variability (zero-crossing rate)

Loudness variability is measured by the zero-crossing rate, which counts the number of times that the voice signal crosses the zero line within each time frame. The zero-crossing rate is calculated as the sum of consecutive zero-crossings divided by the total number of time frames:

$$Zero - crossing\ rates = \frac{1}{T-1} \sum_{\tau=1}^{T} sign(x_\tau) * sign(x_{\tau-1}) \tag{S2}$$

where *x* represents the signal's amplitude, while *T* denotes the number of time frames in the signal.

Pitch variability (F0 standard deviation)

Pitch variability refers to the deviation of the fundamental frequency's standard deviation (Banai et al., 2017).

Talking duration (voicing probability)

Talking duration refers to the duration of speech in a video and is quantified by the voicing probability, which is the probability that a speaker produces speech within a unit time frame. The voicing probability is obtained by normalizing the maximum value of the autocorrelation function (ACF) by the 0th ACF coefficient (Eyben 2015):

$$Voicing\ probability = \frac{ACF_{Max}}{ACF_0} \tag{S3}$$

The variable we refer to as "talking duration", measures the density and verbosity of an influencer's speech, and is equivalent to the voicing probability as defined in previous studies.

OA 2.B. VOCAL CHARACTERISTICS STATISTICS

We measure four prosodic features (average loudness, average pitch, loudness variability, and pitch variability) and one voice quality feature (talking duration). Pitch variability is calculated as the standard deviation of pitch, resulting in four primary features. For each feature, we employ OpenSMILE to capture the 19 statistics described in Table S6, as well as the first-order derivative of each statistic. Consequently, we obtain a total of 152 vocal feature statistics (38 statistics per vocal feature). Table S6 provides functional descriptions of the 19 statistics used in our analysis, which are mainly adopted from Eyben et al. (2010).

Table S6. List of vocal characteristics statistics

Function	Description
Max	Maximum
Min	Minimum
Range	Range
MaxPos	Absolute frame position of the maximum feature value
MinPos	Absolute frame position of the minimum feature value
Mean	Arithmetic mean
Linregc1	Slope of a linear approximation of the voice feature contour
Linregc2	Offset of a linear approximation of the voice feature contour
LinregerrA	Linear error of the computed difference between the linear approximation and voice feature contour
LinregerrQ	Quadratic error of the computed difference between the linear approximation and voice feature contour
Stddev	Standard deviation

Skewness	3rd order moment
Kurtosis	4th order moment
Quantile1	1st quantile (25%)
Quantile2	2nd quantile (50%)
Quantile3	3rd quantile (75%)
Iqr1.2	Interquartile range between the 1st and 2nd quantiles
Iqr1.3	Interquartile range between the 1st and 3rd quantiles
Iqr2.3	Interquartile range between the 2nd and 3rd quantiles

OA 2.C. MAKEUP COLOR ATTRIBUTES

We selected makeup style classes from Alashkar et al. (2017) after conducting a thorough investigation of computer vision literature and commercial online/mobile software applications that create virtual makeup. Taleb Alashkar, the first author, is currently the CTO & Co-Founder of AlgoFace, a virtual makeup software company that develops AI/AR computer vision applications that enable identity-free face Augmented Reality (AR) experiences. To control for makeup style, we incorporated seven color attributes, which are defined in Table S7.

Table S7. Makeup Color Attribute Classes and Model Implementations Summary

No.	Makeup Item	Makeup Attribute	Operationalization	Facial Landmark(s)
1	Foundation	Foundation color	Color of skin representation	3 areas between points (57,8), (39,48), (42,54)
2	Blush	Blush color	Color difference between blushed and non-blushed cheek areas	5 areas between points (36,5)
3	Lip	Lipstick color	Color of lip vermillion	Area between points (55,59)
4		Lip liner	Color difference between the colors of lip vermillion and border	Area between points (50,51)
5	Eyes	Eyeshadow color	Color of the upper eyelid	5 areas between points (19,37)
6		Eyeliner color	Color of the eye border	Point 37
7		Eyebrow color	Color of eyebrows	Point 19

We employ Dlib's 68-point facial landmark detector in the OpenCV package to identify 68 facial landmarks and extract the seven makeup color attributes listed in Table S7. The RGB color values in each area/point specified in Table S7 are then identified, as shown in Figure S1.

The makeup color extraction procedure generated 17 color attributes, with G (Green) value defining each color class, including foundation, blush, lip color, lip liner, eyeshadow, eyeliner, and eyebrow. Due to multicollinearity, R and B values were excluded. Summary statistics are presented in Table S8.

Figure S1. 68 Facial Landmarks

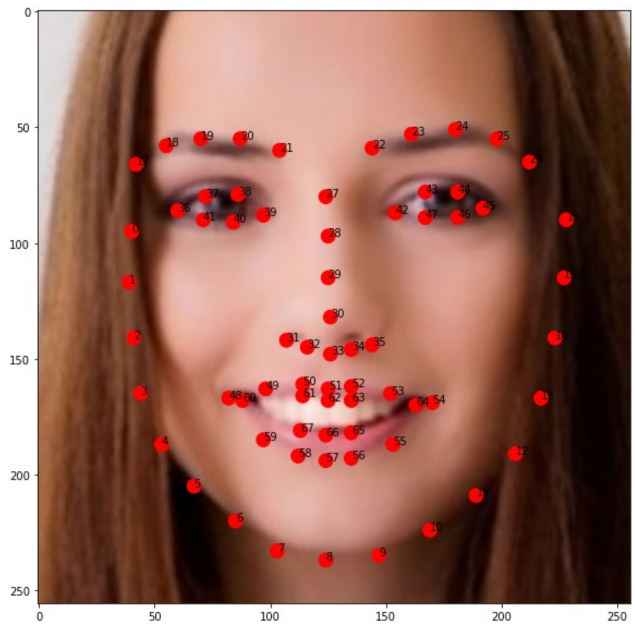


Figure S2. Extraction of Makeup Colors using OpenCV Python Package

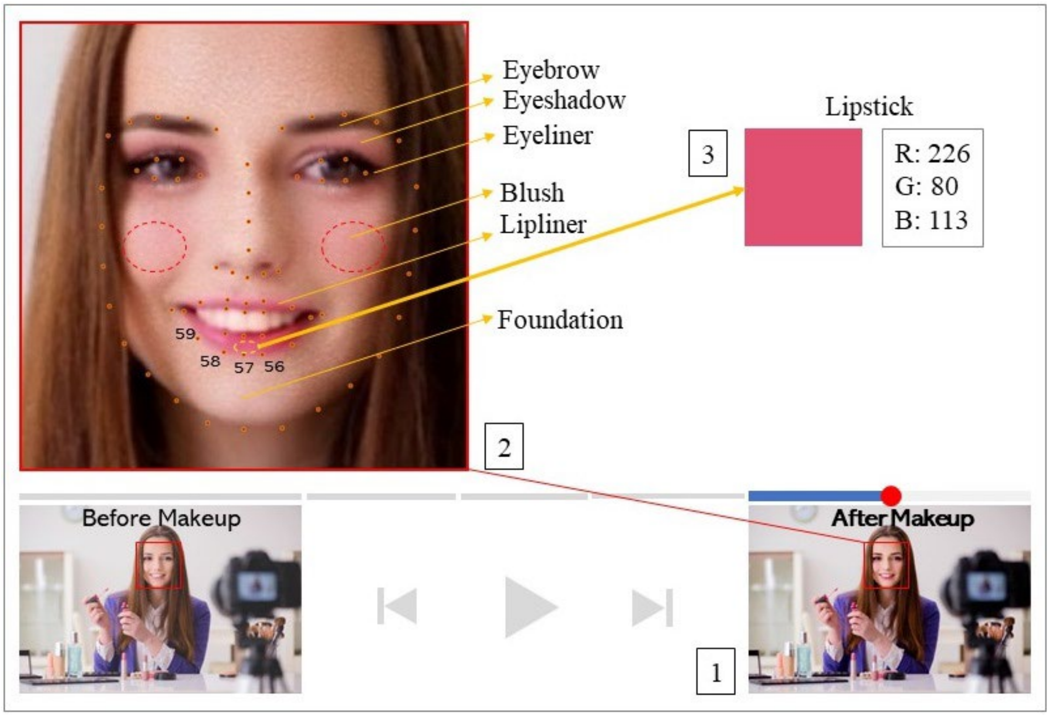


Table S8. Summary Statistics for Makeup Color Attributes

Attribute Name	N	Mean	SD	Min.	Max.
Foundation1 (G)	32,920	0.539	0.241	0.00	1.00
Foundation2 (G)	32,920	0.518	0.231	0.00	1.00
Foundation3 (G)	32,920	0.518	0.232	0.00	1.00
Blush1 (G)	32,920	0.453	0.218	0.00	1.00
Blush2 (G)	32,920	0.456	0.214	0.00	1.00
Blush3 (G)	32,920	0.482	0.218	0.00	1.00
Blush4 (G)	32,920	0.511	0.225	0.00	1.00
Blush5 (G)	32,920	0.531	0.231	0.00	1.00
Lip (G)	32,920	0.379	0.202	0.00	1.00
Lipliner (G)	32,920	0.423	0.197	0.00	1.00
Eyeshadow1 (G)	32,920	0.431	0.225	0.00	1.00
Eyeshadow2 (G)	32,920	0.446	0.218	0.00	1.00
Eyeshadow3 (G)	32,920	0.420	0.209	0.00	1.00
Eyeshadow4 (G)	32,920	0.398	0.204	0.00	1.00
Eyeshadow5 (G)	32,920	0.383	0.201	0.00	1.00
Eyeliner (G)	32,920	0.142	0.116	0.00	1.00
Eyebrow (G)	32,920	0.375	0.197	0.00	1.00

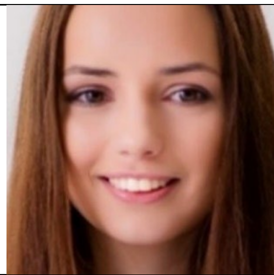
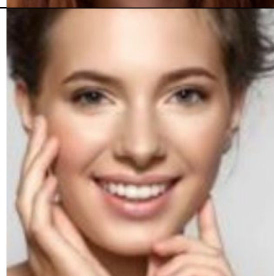
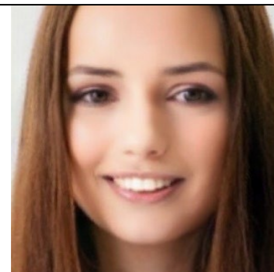
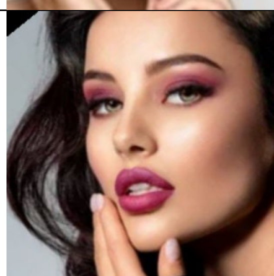
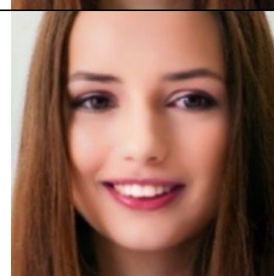
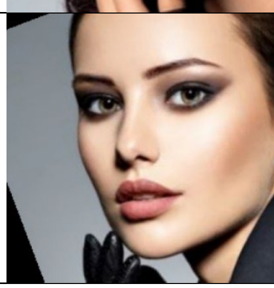
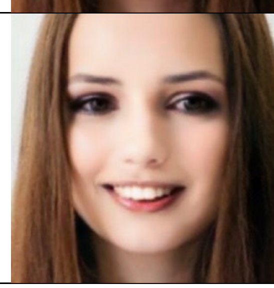
OA 2.D. MAKEUP SPECTRUM

The makeup spectrum quantifies the heaviness of an influencer's makeup by measuring the distance between their original facial image and a makeup-transferred image. This involves two steps: (1) applying a makeup transfer to the influencer's facial image as the baseline, and (2) measuring the cosine distance between the original and the makeup-transferred image. A larger distance signifies a substantial change from the baseline image and indicates that the influencer was wearing relatively heavy makeup.

In step one of our methodology, we utilize BeautyGAN (Li et al. 2018), which adopts the structure of Cycle-GAN (Zhu et al., 2017), an extended Generative Adversarial Network (GAN) that uses adversarial loss and cycle consistency loss. BeautyGAN comprises one generator G and two discriminators: D_A, D_B. The generator's loss function consists of four types of loss: adversarial loss, cycle consistency loss, perceptual loss (based on the differences between high-level features, measured with 16-layer VGG networks pre-trained on the imageNet dataset), and makeup contain loss (local histogram losses through face parsing for makeup applied areas—eyes, face, lips). The discriminators solely contain adversarial loss. Adversarial learning is employed to perform makeup transfer on the baseline image, and Figure S3 demonstrates a baseline face image (top left), three reference makeup images, and three corresponding makeup-transferred images.

Figure S3. Three Makeup Transfer Examples using BeautyGAN

	Reference Image	Makeup Transferred Image

Baseline Image		
Video A		
Video B		
Video C		

In the second step, we compute the cosine distance between the baseline facial image and each of the makeup-transferred faces generated by BeautyGAN, as depicted in Figure S4. The summary statistics of the makeup spectrum can be found in Table S9.

Figure S4. Cosine Distances Between a Baseline Images and Three Makeup-Transferred Images

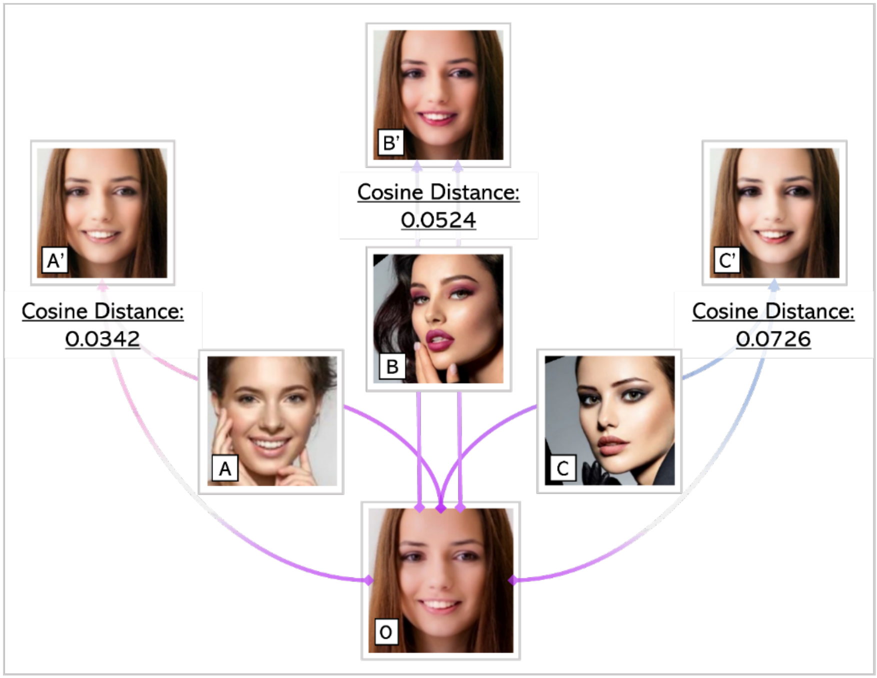


Table S9. Summary Statistics of Makeup Spectrum (Cosine Distance Score)

Variable	N	Mean	SD	Min.	Max.
Makeup heaviness	32,920	0.056	0.038	0	0.282

OA 2.E. NEURAL IMAGE ASSESSMENT (NIMA) SCORE

The Neural Image Assessment (NIMA) score employs advanced deep object recognition networks, such as VGG16, Inception-V2, and MobileNet, to predict image quality and aesthetics, as proposed by Talebi and Milanfar (2018). The NIMA score has gained widespread recognition for its ability to predict consumers' aesthetic perception of images. We calculated the mean and standard deviation of the predicted distribution of aesthetic ratings ranging from 1 to 10 provided by the NIMA models. Among the three benchmark models used in Talebi and Milanfar (2018), we chose MobileNet, a deep convolutional neural network model, for its superior efficiency and comparable performance.

As a demonstration of the NIMA score's efficacy, we present four sample photos with low and high NIMA scores, sourced from the AVA dataset used by Talebi and Milanfar (2018) and our influencer video dataset. The examples indicate that images with low NIMA scores are typically perceived as less visually appealing compared to those with high NIMA scores. Summary statistics are presented in Table S10.

Figure S5. NIMA Score and Images

Original AVA dataset from NIMA		Influencer video datasets	
3.55	6.38	3.89	6.01

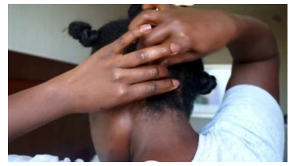
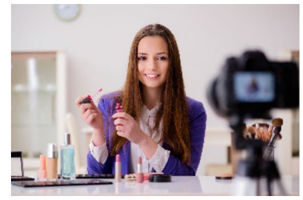
			

Table S10. Summary Statistics of Neural Image Assessment Score (Score from 1–10)

Variable	N	Mean	SD	Min.	Max.
NIMA Score	32,920	5.433	0.55	3.266	6.752

OA 2.F. FACIAL ATTRACTIVENESS

The influence of higher attractiveness on eliciting positive responses, known as the "attractiveness premium," may introduce bias in our estimation if influencers appear more attractive in sponsored videos than non-sponsored ones. To mitigate this potential bias, we develop a CNN model trained on the SCUT-FBP5500 dataset (Liang et al., 2018) to assign an attractiveness score (1 to 5) to each influencer. Summary statistics are presented in Table S11.

Table S11. Summary Statistics of Facial Attractiveness (Score from 1–5)

Variable	N	Mean	SD	Min.	Max.
Attractiveness	32,920	2.907	0.363	1.854	4.755

OA 2.G. DEMOGRAPHICS & EMOTIONS

DeepFace is utilized to extract key facial features of the influencers including their age, race, and emotions, following the methodology used by Mittal et al. (2020). The probabilities of seven emotions and six races, as presented in Table S12, are computed to analyze the data.

Table S12. Summary Statistics of Demographics, Facial Area, and Emotions

Variable	N	Mean	SD	Min.	Max.
Age	32,920	30.529	3.490	19	56.000
Facial area	32,920	10495.419	7864.187	1296	88506.000
Emotions					
Angry (%)	32,920	0.060	0.153	0	1
Disgust (%)	32,920	0.013	0.079	0	1
Fear (%)	32,920	0.100	0.190	0	1
Happy (%)	32,920	0.287	0.369	0	1

Neutral (%)	32,920	0.329	0.363	0	1
Sad (%)	32,920	0.163	0.258	0	1
Surprise (%)	32,920	0.049	0.168	0	1
Races					
Asian (%)	32,920	0.155	0.263	0	1
Black (%)	32,920	0.164	0.322	0	1
Indian (%)	32,920	0.060	0.094	0	1
Latino/Hispanic (%)	32,920	0.159	0.127	0	0.818
Middle Eastern (%)	32,920	0.126	0.142	0	0.988
White (%)	32,920	0.335	0.321	0	1

OA 2.H. IMAGE OBJECT LABELS

We employed Mask R-CNN (He et al., 2017) for object identification in a more sophisticated manner. To train the Mask R-CNN model, we used the iMaterialist fine-grained visual segmentation training dataset, which is a fashion taxonomy created by fashion domain experts and crowdsourced workers, containing 46 apparel objects (27 main apparel items and 19 apparel parts) and 92 related fine-grained attributes.[8] The dataset comprises 45,195 images and 331,213 segmentations, which we divided into training (80%; 36,156 images and 264,949 segments) and validation (20%; 9,039 images and 66,264 segments) sets. We trained eight models and selected the best-performing one, with a learning rate of 0.002 in the first two models, 0.001 in the second two, and 0.00002 in the last four. We used most of the original parameters of the Mask R-CNN model provided by He et al. (2017), which are associated with the FPN and RPN networks (e.g., RPN_ANCHOR_RATIOS).

Comparing the performances of the models on three metrics, namely Loss, MRCNN Class Loss, and MRCNN Mask Loss, Figure S6 indicates that the 8th model (model ID = 7) outperforms the others. Therefore, we use it for further analyses. Figure S7 provides examples of the object detection outcomes utilizing the 8th model. The model performs remarkably well in detecting objects such as 'sleeves' and 'shoes.' In Table S13, we display the summary statistics of detected image objects.

Figure S6. Performance Comparison of 8 Mask R-CNN Models

Train loss vs. Validation loss	MRCNN class loss vs. validation MRCNN class loss	MRCNN mask loss vs. validation MRCNN mask loss

[8] It is part of the FGVC7 workshop at CVPR. For more information, please visit https://sites.google.com/view/fgvc7

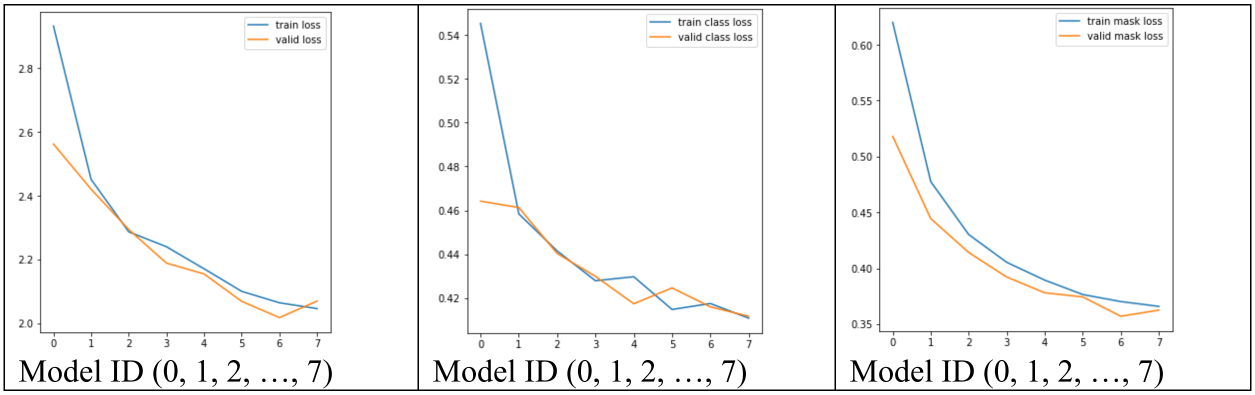


Figure S7. Three Examples of Image Labels From Mask R-CNN

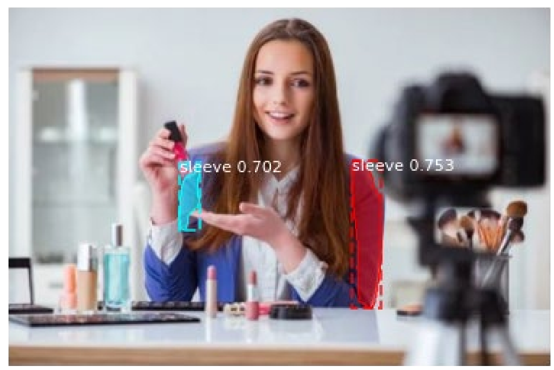
Raw Image	Image with Mask R-CNN Object Labels
	
	

Table S13. Summary Statistics of Detected Image Objects

Variable	N	Mean	SD	Min.	Max.
Top, t-shirt, sweatshirt	32,920	0.016	0.127	0	1
Jacket	32,920	0.000	0.000	0	1
Pants	32,920	0.004	0.062	0	1
Skirt	32,920	0.000	0.011	0	1

Dress	32,920	0.172	0.378	0	1
Glasses	32,920	0.005	0.072	0	1
Hat	32,920	0.000	0.012	0	1
Watch	32,920	0.000	0.006	0	1
Belt	32,920	0.000	0.010	0	1
Tights, stockings	32,920	0.000	0.022	0	1
Shoe	32,920	0.086	0.280	0	1
Bag, wallet	32,920	0.000	0.021	0	1
Collar	32,920	0.000	0.022	0	1
Lapel	32,920	0.000	0.021	0	1
Sleeve	32,920	0.657	0.475	0	1
Pocket	32,920	0.001	0.027	0	1
Neckline	32,920	0.241	0.427	0	1

OA 2.I. BERTOPIC VERBAL TOPIC MODELING

We utilize the BERTopic model to extract topics from the narrative in the videos. BERTopic is a powerful topic modeling technique that utilizes transformers and c-TF-IDF to create dense clusters that enable the formation of coherent and interpretable topics that capture the most salient words in the topic description (Grootendorst 2020). The general model implementation involves using the pre-trained algorithm developed by Grootendorst (2020).

Using a pre-trained approach, we employ two sentence transformers, namely, (1) paraphrase-MiniLM-L6-v2 and (2) paraphrase-multilingual-MiniLM-L12-v2, to analyze the semantic similarity of the text. While the former is an English BERT-based model trained for semantic similarity tasks, the latter works for over 50 languages. Further, we reduce the dimensionality of the text using the UMAP technique and cluster them using HDBSCAN. Finally, we use class-based TF-IDF to create topic representations from the clusters. A comprehensive overview of the verbal topics is presented in Table S14.

Table S14. Summary Statistics of Verbal Topics

Variable	N	Mean	SD	Min.	Max.
Verbal topic 0	32,920	0.105	0.307	0	1
Verbal topic 1	32,920	0.013	0.113	0	1
Verbal topic 2	32,920	0.008	0.089	0	1
Verbal topic 3	32,920	0.018	0.133	0	1
Verbal topic 4	32,920	0.016	0.125	0	1
Verbal topic 5	32,920	0.009	0.095	0	1
Verbal topic 6	32,920	0.006	0.076	0	1
Verbal topic 7	32,920	0.007	0.084	0	1
Verbal topic 8	32,920	0.006	0.078	0	1
Verbal topic 9	32,920	0.005	0.072	0	1

OA 2.J. VERBAL SOPHISTICATION MEASURES

We employ the use of TAALES, a publicly available lexical sophistication toolkit (Kyle and Crossley 2014; Kyle et al. 2018), to quantify the level of language sophistication present in our data. TAALES 2.0 is a powerful tool for the automatic analysis of lexical sophistication, offering a range of lexical categories, including unigram frequency, n-gram frequency, and psycholinguistic word properties (Kyle et al. 2018). By utilizing this software, we are able to obtain a more nuanced understanding of the linguistic characteristics present in our dataset.

The lexical sophistication toolkit TAALES (Kyle and Crossley 2014; Kyle et al. 2018) is constructed by amalgamating several text corpora from the existing literature. To measure word frequency and range, TAALES 2.0 integrates Brown corpus, Lorge's corpus of popular magazine articles, London-Lund corpus of conversation, SUBTLEXus corpus, British Natural Corpus (BNC), Corpus of Contemporary American English (COCA), and Hyperspace Analogue to Language (HAL). N-gram frequency and range are derived from BNC and COCA corpora. In terms of psycholinguistic properties, the package employs the MRC corpus, the age of acquisition norms, and the concreteness norms using the Brysbaert model (Brysbaert et al. 2014).

We adopt a comprehensive approach that leverages both corpus-based lexical measures and psycholinguistic norm measures to measure lexical sophistication. Specifically, for the lexical sophistication measures, we utilize the frequencies of words (unigrams), bigrams, and trigrams in the Corpus of Contemporary American English (COCA), which is the most recent and genre-balanced corpus of American English. COCA contains over 1.1+ billion words of text from a variety of genres, including spoken, fiction, popular magazines, newspapers, academic texts, and web pages. To provide a richer picture of lexical sophistication, we include both bigram and trigram alongside unigram. In addition to corpus-based measures, we incorporate psycholinguistic norm measures using the Brysbaert Concreteness dataset (Brysbaert et al. 2014), which comprises 37,058 English words and 2,896 bigrams, with ratings obtained from over 4,000 crowdsourced participants in a norming study. The resulting summary of the lexical verbal sophistication measures is presented in Table S15, with the summary statistics (semi-log-transformed) provided in Table S16.

Table S15. Selected Verbal Sophistication Measures

Verbal sophistication feature name	Type of words	Category
COCA spoken word frequency	Content words (CW)	Word frequency
COCA spoken bigram frequency	Content words (CW)	N-gram frequency
COCA spoken trigram frequency	Content words (CW)	N-gram frequency
Brysbaert concreteness	Content words (CW)	Psycholinguistic norms

Table S16. Summary Statistics of Lexical Sophistication Features

Verbal sophistication feature name	N	Mean	SD	Min.	Max.

COCA spoken unigram frequency	32,920	1381.826	208.530	0.00	3821.903
COCA spoken bigram frequency	32,920	237.798	39.519	0.00	792.609
COCA spoken trigram frequency	32,920	24.362	7.035	0.00	180.258
Brysbaert concreteness	32,920	2.904	0.155	0.00	4.300

OA 2.K. THE 30 MOST-FEATURED BRANDS & PRODUCT TYPES

To mitigate any potential endogeneity issues arising from the usage of a product of a specific brand in videos, we incorporate indicators for the top 30 brands with the highest sponsorship occurrences during our sample period. To provide insight into the characteristics of the brands and products in our dataset, we present Table S17, which provides the summary statistics of the 30 most prominent brands, and Table S18, which provides the summary statistics of the 30 most frequently featured product types.

Table S17. Summary Statistics of 30 Most-Featured Brands

Brand name	N	Mean	SD	Min.	Max.
mac cosmetics	32,920	0.124	0.329	0	1
anastasia beverly hills	32,920	0.087	0.282	0	1
nyx cosmetics	32,920	0.078	0.268	0	1
maybelline	32,920	0.081	0.273	0	1
too faced	32,920	0.090	0.287	0	1
urban decay	32,920	0.087	0.282	0	1
tarte	32,920	0.085	0.279	0	1
morphe	32,920	0.077	0.266	0	1
benefit cosmetics	32,920	0.079	0.269	0	1
ofra	32,920	0.088	0.283	0	1
loreal	32,920	0.069	0.254	0	1
nars	32,920	0.069	0.254	0	1
colourpop cosmetics	32,920	0.052	0.222	0	1
becca cosmetics	32,920	0.047	0.212	0	1
laura mercier	32,920	0.052	0.222	0	1
makeup forever	32,920	0.049	0.215	0	1
kat von d	32,920	0.038	0.191	0	1
it cosmetics	32,920	0.049	0.215	0	1
smashbox	32,920	0.045	0.207	0	1
elf cosmetics	32,920	0.034	0.181	0	1
wet n wild	32,920	0.027	0.163	0	1
hourglass	32,920	0.035	0.184	0	1
marc jacobs	32,920	0.033	0.179	0	1
milani	32,920	0.023	0.149	0	1

rimmel london	32,920	0.026	0.158	0	1
covergirl	32,920	0.029	0.167	0	1
la girl	32,920	0.028	0.164	0	1
beauty blender	32,920	0.028	0.164	0	1
huda beauty	32,920	0.026	0.158	0	1
jouer	32,920	0.031	0.173	0	1

Table S18. Summary Statistics of 30 Most-Featured Product Types

Product type	N	Mean	SD	Min.	Max.
eye other	32,920	0.311	0.463	0	1
foundation	32,920	0.258	0.438	0	1
lash	32,920	0.250	0.433	0	1
brow	32,920	0.244	0.430	0	1
brush	32,920	0.246	0.431	0	1
palette	32,920	0.239	0.427	0	1
powder	32,920	0.222	0.416	0	1
tool	32,920	0.188	0.391	0	1
conceal	32,920	0.204	0.403	0	1
liner	32,920	0.189	0.391	0	1
lip stick	32,920	0.180	0.384	0	1
shadow	32,920	0.166	0.372	0	1
highlight	32,920	0.171	0.377	0	1
blush	32,920	0.167	0.373	0	1
mascara	32,920	0.173	0.378	0	1
bronz	32,920	0.144	0.351	0	1
primer	32,920	0.136	0.343	0	1
lip other	32,920	0.119	0.324	0	1
contour	32,920	0.110	0.313	0	1
oil	32,920	0.120	0.325	0	1
spray	32,920	0.094	0.292	0	1
mask	32,920	0.074	0.261	0	1
moisturi	32,920	0.051	0.221	0	1
cleans	32,920	0.054	0.226	0	1
serum	32,920	0.047	0.211	0	1
lip gloss	32,920	0.040	0.195	0	1
mist	32,920	0.042	0.200	0	1
comb	32,920	0.037	0.189	0	1
fac other	32,920	0.030	0.169	0	1
spong	32,920	0.030	0.171	0	1

OA 2.L. VIDEO CHARACTERISTICS & POPULARITY MEASURES

Table S19 displays the summary statistics for seven distinct video characteristics and six measures of popularity.

Table S19. Summary Statistics of Video Characteristics and Popularity Measures

	N	Mean	SD	Min.	Max.
Video Characteristics					
Duration (Seconds)	32,920	676.877	405.939	13.000	9,466
View count	32,920	111,961.013	411,158.807	171.000	23,889,090
Like count	32,920	3,763.070	8,788.897	6.000	305,269
Dislike count	32,920	96.613	368.660	0.000	19,376
Comment count	32,920	343.316	1,548.491	1.000	102,655
Title length (number of character)	32,920	52.011	18.225	4.000	115
Description length (number of character)	32,920	1,850.337	1,033.829	0.000	5,697
Tenure (number of days)	32,920	1,522.156	788.944	0.000	4,302
Popularity measure					
Follower count (raw)	32,920	416576.194	595425.023	804	3,639,722.96
Total views (raw)	32,920	87447094.622	17630255.56	209,551	99,868,273
IG follower count (raw)	32,920	220360.461	460080.574	33.226	3,112,907
Follower count (diff.)	32,920	295.719	543.151	-76.990	3,526.947
Total views (diff.)	32,920	27366.140	23827.554	-8373.408	190,129.5
IG follower count (diff.)	32,920	77.689	373.013	-48.783	2,759.152

OA 2.M. SIX VIDEO FORMATS

Table S20. Summary Statistics of the Six Video Formats

Video Format (by video)	N	Mean	SD	Min.	Max.
GRWM (dummy)	32,920	0.183	0.387	0	1
Haul (dummy)	32,920	0.389	0.487	0	1
Review (dummy)	32,920	0.317	0.465	0	1
Routine (dummy)	32,920	0.287	0.453	0	1
Tutorial (dummy)	32,920	0.696	0.460	0	1
Vlog (dummy)	32,920	0.410	0.492	0	1

OA SECTION 3. PROPENSITY SCORE MATCHING (PSM)-IV APPROACH

In this study, we employ the propensity score matching with instrumental variables (PSM-IV) approach to match sponsored videos with non-sponsored videos based on pre-treatment observables. We aim to address potential confounding effects on the treatment effect using the TSCS data matching method proposed by Imai et al. (2021). Specifically, we match on two broad categories of pre-treatment observables: sponsorship treatment history and influencer/video characteristics. The former includes the number of videos, sponsored videos, and sponsored videos by product category, while the latter includes various features, such as influencer and month fixed effects, visual features (e.g., age, emotions, facial attractiveness), verbal topics and sophistication measures, popularity measures, featured brands and product types, video formats, monthly advertising spending, and various video statistics. By matching on these pre-treatment observables, we aim to minimize the differences between the sponsored and non-sponsored videos and estimate the causal effect of sponsorship on consumer sentiment.

As introduced in Equation (S4), we utilize a logistic regression to estimate the propensity score by employing the binary sponsorship indicator as the dependent variable and considering the observable characteristics.

$$\begin{aligned}
Sponsorship_{ijt} &= \beta_0 + \overrightarrow{\beta_1} \times \overrightarrow{TreatmentHistory_{i,t-Q}} + \overrightarrow{\alpha_0} \times \overrightarrow{VisualFeatures_{i,t-Q}} + \overrightarrow{\alpha_1} \times \overrightarrow{VerbalFeatures_{i,t-Q}} + \overrightarrow{\alpha_2} \times \\
&\overrightarrow{TextualFeatures_{i,t-Q}} + \overrightarrow{\alpha_3} \times \overrightarrow{Popularity_{i,t-Q}} + \overrightarrow{\alpha_4} \times \overrightarrow{Brands_{i,t-Q}} + \overrightarrow{\alpha_5} \times \\
&\overrightarrow{Types_{i,t-Q}} + \alpha_6 \times AdSpending_{i,t-Q} + I_i + \tau_t + \epsilon_{ijt},\ \epsilon_{ijt} \sim N(0, \sigma_\epsilon^2)
\end{aligned}$$

$$\begin{aligned}
\overrightarrow{TreatmentHistory_{i,t-Q}} &= \left[Video_{t-Q}, Sponsorship_{t-Q}, SponsorshipCategory_{t-Q}\right]' \qquad \text{(S4)} \\
\overrightarrow{VisualFeatures_{i,t-Q}} &= [\overrightarrow{MakeupStyle_{i,t-Q}}, MakeupSpectrum_{i,t-Q}, \\
&\quad Attractiveness_{i,t-Q}, Age_{i,t-Q}, \overrightarrow{Race_{i,t-Q}}, \overrightarrow{Objects_{i,t-Q}}]' \\
\overrightarrow{VerbalFeatures_{i,t-Q}} &= \left[\overrightarrow{Topics_{i,t-Q}}, \overrightarrow{Sophistications_{i,t-Q}}\right]' \\
\overrightarrow{TextualFeatures_{i,t-Q}} &= [Duration_{i,t-Q}, Views_{i,t-Q}, Likes_{i,t-Q}, Dislikes_{i,t-Q}, Comments_{i,t-Q}, \\
&\quad TitleLength_{i,t-Q}, DescLength_{i,t-Q}, Tenure_{i,t-Q}, \overrightarrow{Formats_{i,t-Q}}]
\end{aligned}$$

$$\overrightarrow{Popularity_{i,t-Q}}$$
$$= [Followers_{i,t-Q}, AccountViews_{i,t-Q}, InstagramFollowers_{i,t-Q},$$
$$\Delta Followers_{i,t-Q}, \Delta AccountViews_{i,t-Q}, \Delta InstagramFollowers_{i,t-Q}]'$$

Upon computing the propensity score for each video, we perform a matching procedure to pair each sponsored video with a non-sponsored video that exhibits a similar propensity score. This approach enables us to compare equivalent videos that differ only in sponsorship status. We then employ Equation (S5) to investigate whether influencers alter any of their five vocal characteristics in sponsored videos compared to their equivalent non-sponsored counterparts. This methodology allows us to answer the first research question concerning the effect of sponsorship on vocal characteristics.

$$Voice_{ijt}^{k} = \beta_0 + \beta_1 \times Sponsorship_{ijt} + \vec{\Gamma} \times \overrightarrow{ControlVariables}_{ijt} + I_i + \tau_t + \epsilon_{ijt},\ \epsilon_{ijt} \sim N(0, \sigma_\epsilon^2) \quad \text{(S5)}$$

In Equation (S5), the vector $\vec{\Gamma}$ comprises the coefficients of content variables, denoted by α_n, where n takes on values from 1 to 144. Additionally, the vector $\overrightarrow{Control\ variables}$ incorporates 144 control variables, namely, 17 makeup colors, 17 image objects, makeup heaviness, facial attractiveness, NIMA image aesthetic score, age, 6 emotions, 5 races, 10 verbal narrative topics, 4 verbal sophistication measures, 30 most frequently appearing brands, 30 most frequently appearing product types, 6 video formats, monthly ad spending, duration, view count, like count, dislike count, comment count, title length, description length, tenure, the raw number of followers, total views, and Instagram followers, the average difference in the number of followers, total views, and Instagram followers, year, month, and influencer fixed effects.

To address the possible endogeneity of sponsorship and answer the second research question, we adopt the IV approach. Specifically, we estimate the first-stage regression for the sponsorship indicator using Equation (S6), and the first-stage regression for the interaction between the sponsorship indicator and vocal characteristics using Equation (S7).

$$Sponsorship_{ijt} = \beta_0 + \overrightarrow{\beta_1} \times \overrightarrow{Concurrent\ videos\ (m, m')} + \overrightarrow{\beta_2} \times \overrightarrow{Voice_{ijt}^{k}} \times \overrightarrow{Concurrent\ videos\ (m, m')} + \vec{\Gamma} \times \overrightarrow{ControlVariables} + I_i + \tau_t + \epsilon_{ijt} \quad \text{(S6)}$$

$$Sponsored\ voice\ characteristics_{ijt} = \beta_0 + \overrightarrow{\beta_1} \times \overrightarrow{Concurrent\ videos\ (m, m')} + \overrightarrow{\beta_2} \times \overrightarrow{Voice_{ijt}^{k}} \times \overrightarrow{Concurrent\ videos\ (m, m')} + \quad \text{(S7)}$$

$$\overrightarrow{\Gamma} \times \overrightarrow{ControlVariables} + I_i + \tau_t + \epsilon_{ijt} \text{ where}$$

$$Sponsored\ voice\ characteristics_{ijt} = Sponsorship_{ijt} \times Voice_{ijt}^{k} \text{ and}$$

$$Voice_{ijt}^{k} \in \{\overline{averageLoudness_{ijt}}, \overline{averagePitch_{ijt}}, \overline{loudnessVariability_{ijt}}, \overline{pitchVariability_{ijt}}, \overline{talkingDuration_{ijt}}\}$$

Drawing inspiration from Papies et al. (2017), we estimate the effect of sponsorship on vocal characteristics by regressing the interaction between sponsorship and vocal characteristics on the covariates (same as in Equation (S6)). Readers may refer to section 18.6.1.2 in Papies et al. (2017) for further details.

To address potential endogeneity of sponsorship in our study, we employ an instrumental variable (IV) approach. Specifically, we use the instrumented sponsorship indicator and the instrumented interaction between sponsorship and vocal characteristics in the second-stage regression. The instrumental variables are obtained from the first-stage regressions (Equations S6 and S7), which utilize the same independent variables as in Equation (S5).

$$Sentiments_{ijt} = \beta_0 + \beta_1 \times \widetilde{Sponsorship}_{ijt} + \beta_2 \times \overrightarrow{Voice_{ijt}^{k}} + \beta_3 \times \overrightarrow{Sponsored\ voice\ \widetilde{characteristics}_{ijt}} + \overrightarrow{\Gamma} \times \overrightarrow{Contents\ variables} + \epsilon_{ijt} \tag{S8}$$

where $\widetilde{Sponsorship}_{ijt}$ represents the fitted value obtained after performing the first-stage regression analysis as described in Equation (S6). In addition, the vector $\overrightarrow{Sponsored\ voice\ \widetilde{characteristics}_{ijt}}$ denotes the set of fitted values obtained from the five individual first-stage regression analyses for the five distinct vocal characteristics examined, namely average loudness, average pitch, loudness variability, pitch variability, and talking duration.

In the subsequent section, we present the outcomes of the Propensity Score Matching (PSM) methodology and verify if the chosen instrument complies with the relevance condition and exclusion restriction.

OA 3.A. PSM REGRESSION RESULTS

We utilize propensity score matching along with a set of 1,016 individual fixed effects and content covariates. To evaluate the effectiveness of the matching procedure, Table S21 presents the summary statistics of the entire panel sample, as well as the matched sample. It should be noted that individual and time fixed effects are also incorporated in the matching process, although these are not reported in the summary statistics due to space constraints.

Table S21. Summary Statistics of the All & Propensity-Score Matched Data

	All Data			Propensity-Score Matched Data		
Matching Variable	Means Treated	Means Control	Std. Mean Diff.	Means Treated	Means Control	Std. Mean Diff.
Distance	0.2486	0.1385	0.7627	0.2486	0.2421	0.0447
Pre-treated Sponsorship History						
Number of videos	2.7998	2.8433	-0.0655	2.7998	2.8024	-0.0038
Number of sponsored videos	1.2472	0.8444	0.5297	1.2472	1.2358	0.0150
Number of sponsored fashion videos	0.3137	0.1875	0.2445	0.3137	0.3191	-0.0104
Number of sponsored hair videos	0.2345	0.1200	0.2335	0.2345	0.2238	0.0219
Number of sponsored makeup videos	0.5720	0.4217	0.2415	0.5720	0.5731	-0.0019
Number of sponsored other videos	0.4348	0.2745	0.2635	0.4348	0.4341	0.0011
Number of sponsored skincare videos	0.1594	0.1152	0.1322	0.1594	0.1610	-0.0045
Influencer & Video Characteristics						
Duration (Seconds)	6.4443	6.4263	0.0330	6.4443	6.4474	-0.0057
View count	10.8477	10.7611	0.0590	10.8477	10.8566	-0.0061
Like count	7.4861	7.4188	0.0501	7.4861	7.4998	-0.0101
Dislike count	3.8772	3.8173	0.0492	3.8772	3.8897	-0.0102
Comment count	5.0748	5.0961	-0.0173	5.0748	5.0924	-0.0143
Title length	3.9250	3.8975	0.0816	3.9250	3.9227	0.0068
Description length	7.3484	7.2868	0.0905	7.3484	7.3486	-0.0002
Tenure (number of days)	7.0701	6.8515	0.2586	7.0701	7.0869	-0.0198
Follower count (raw)	12.1222	11.7704	0.2449	12.1222	12.1413	-0.0133
Total views (raw)	18.0226	17.7788	0.1694	18.0226	18.0115	0.0077
Instagram follower count (raw)	10.7347	10.5547	0.0844	10.7347	10.7623	-0.0130
Follower count (diff.)	-0.0157	0.0029	-0.0202	-0.0157	-0.0064	-0.0101
Total views (diff.)	-0.0697	0.0128	-0.0856	-0.0697	-0.0588	-0.0113
Instagram follower count (diff.)	0.0155	-0.0029	0.0187	0.0155	0.0171	-0.0017
Format - GRWM (dummy)	0.1713	0.1463	0.0995	0.1713	0.1733	-0.0078
Format - haul (dummy)	0.3766	0.3537	0.0662	0.3766	0.3785	-0.0053
Format - review (dummy)	0.3169	0.3484	-0.0917	0.3169	0.3162	0.0021
Format - routine (dummy)	0.2627	0.2549	0.0263	0.2627	0.2659	-0.0105
Format - tutorial (dummy)	0.6923	0.7024	-0.0320	0.6923	0.6923	0.0001

Format - vlog (dummy)	0.4129	0.3886	0.0587	0.4129	0.4124	0.0011
COCA spoken unigram frequency	6.9809	6.8568	0.0926	6.9809	6.9800	0.0007
COCA spoken bigram frequency	5.2770	5.1897	0.0855	5.2770	5.2781	-0.0011
COCA spoken trigram frequency	3.1093	3.0611	0.0779	3.1093	3.1102	-0.0014
Brysbaert's Concreteness	1.3114	1.2877	0.0936	1.3114	1.3113	0.0004
Verbal topic 0	0.0826	0.0605	0.1332	0.0826	0.0762	0.0385
Verbal topic 1	0.0125	0.0155	-0.0769	0.0125	0.0124	0.0016
Verbal topic 2	0.0084	0.0133	-0.1285	0.0084	0.0088	-0.0110
Verbal topic 3	0.0126	0.0097	0.0709	0.0126	0.0127	-0.0028
Verbal topic 4	0.0093	0.0091	0.0061	0.0093	0.0094	-0.0028
Verbal topic 5	0.0081	0.0080	0.0013	0.0081	0.0083	-0.0067
Verbal topic 6	0.0065	0.0083	-0.0763	0.0065	0.0067	-0.0082
Verbal topic 7	0.0068	0.0079	-0.0396	0.0068	0.0069	-0.0041
Verbal topic 8	0.0047	0.0058	-0.0581	0.0047	0.0048	-0.0037
Verbal topic 9	0.0045	0.0052	-0.0337	0.0045	0.0045	-0.0025
Age	3.4361	3.4194	0.0773	3.4361	3.4359	0.0010
Angry (%)	0.0587	0.0615	-0.0434	0.0587	0.0584	0.0048
Disgust (%)	0.0127	0.0129	-0.0086	0.0127	0.0125	0.0057
Fear (%)	0.1014	0.1049	-0.0403	0.1014	0.1012	0.0033
Happy (%)	0.2799	0.2629	0.0959	0.2799	0.2801	-0.0014
Neutral (%)	0.3301	0.3262	0.0231	0.3301	0.3303	-0.0014
Sad (%)	0.1645	0.1725	-0.0690	0.1645	0.1650	-0.0046
Surprise (%)	0.0492	0.0501	-0.0122	0.0492	0.0489	0.0045
Asian (%)	0.1541	0.1426	0.0500	0.1541	0.1538	0.0013
Black (%)	0.1642	0.1235	0.1367	0.1642	0.1611	0.0104
Indian (%)	0.0595	0.0581	0.0195	0.0595	0.0605	-0.0144
Latino/Hispanic (%)	0.1568	0.1576	-0.0081	0.1568	0.1583	-0.0159
Middle Eastern (%)	0.1261	0.1343	-0.0752	0.1261	0.1269	-0.0072
White (%)	0.3357	0.3749	-0.1390	0.3357	0.3357	0.0000
Facial attractiveness	1.3577	1.3549	0.0273	1.3577	1.3580	-0.0038
Makeup heaviness	0.0447	0.0411	0.1301	0.0447	0.0445	0.0076
Top, t-shirt, sweatshirt	0.0128	0.0116	0.0301	0.0128	0.0128	-0.0004
Jacket	0.0000	0.0000	-0.0099	0.0000	0.0000	0.0070
Pants	0.0029	0.0028	0.0022	0.0029	0.0029	-0.0032
Skirt	0.0001	0.0000	0.0110	0.0001	0.0001	-0.0099
Dress	0.1339	0.1332	0.0046	0.1339	0.1335	0.0028
Glasses	0.0034	0.0035	-0.0083	0.0034	0.0035	-0.0064
Hat	0.0001	0.0001	0.0126	0.0001	0.0001	-0.0075

Watch	0.0000	0.0000	-0.0592	0.0000	0.0000	-0.0004
Belt	0.0000	0.0000	0.0021	0.0000	0.0000	0.0041
Tights, stockings	0.0003	0.0004	-0.0228	0.0003	0.0004	-0.0173
Shoe	0.0676	0.0668	0.0090	0.0676	0.0684	-0.0092
Bag, wallet	0.0004	0.0003	0.0179	0.0004	0.0004	0.0122
Collar	0.0004	0.0004	-0.0003	0.0004	0.0004	-0.0065
Lapel	0.0004	0.0004	-0.0013	0.0004	0.0005	-0.0036
Sleeve	0.5245	0.5150	0.0358	0.5245	0.5245	-0.0001
Pocket	0.0006	0.0006	0.0060	0.0006	0.0007	-0.0050
Neckline	0.1897	0.1806	0.0559	0.1897	0.1903	-0.0034
Foundation1 (G)	0.4295	0.4268	0.0133	0.4295	0.4301	-0.0030
Foundation2 (G)	0.4143	0.4117	0.0132	0.4143	0.4146	-0.0017
Foundation3 (G)	0.4153	0.4125	0.0140	0.4153	0.4154	-0.0005
Blush1 (G)	0.3602	0.3614	-0.0065	0.3602	0.3605	-0.0016
Blush2 (G)	0.3629	0.3641	-0.0067	0.3629	0.3632	-0.0015
Blush3 (G)	0.3841	0.3836	0.0028	0.3841	0.3841	0.0001
Blush4 (G)	0.4074	0.4053	0.0110	0.4074	0.4074	-0.0001
Blush5 (G)	0.4236	0.4210	0.0135	0.4236	0.4236	-0.0001
Lip (G)	0.3043	0.2935	0.0724	0.3043	0.3038	0.0035
Lipliner (G)	0.3384	0.3326	0.0366	0.3384	0.3381	0.0023
Eyeshadow1 (G)	0.3438	0.3405	0.0192	0.3438	0.3442	-0.0024
Eyeshadow2 (G)	0.3554	0.3517	0.0214	0.3554	0.3557	-0.0018
Eyeshadow3 (G)	0.3343	0.3297	0.0282	0.3343	0.3344	-0.0007
Eyeshadow4 (G)	0.3171	0.3118	0.0341	0.3171	0.3170	0.0005
Eyeshadow5 (G)	0.3050	0.2998	0.0350	0.3050	0.3050	-0.0000
Eyeliner (G)	0.1124	0.1103	0.0317	0.1124	0.1122	0.0028
Eyebrow (G)	0.2992	0.2992	-0.0004	0.2992	0.2996	-0.0033
NIMA Score	1.8070	1.7789	0.0894	1.8070	1.8085	-0.0048
mac_cosmetics	0.1315	0.1468	-0.0758	0.1315	0.1328	-0.0063
anastasia_beverly_hills	0.0913	0.1172	-0.1662	0.0913	0.0937	-0.0155
nyx_cosmetics	0.0831	0.1013	-0.1195	0.0831	0.0844	-0.0091
maybelline	0.0834	0.1007	-0.1236	0.0834	0.0862	-0.0199
too_faced	0.0918	0.1018	-0.0644	0.0918	0.0935	-0.0109
urban_decay	0.0863	0.1031	-0.1223	0.0863	0.0877	-0.0100
tarte	0.0882	0.1000	-0.0752	0.0882	0.0890	-0.0051
ofra	0.0874	0.0956	-0.0324	0.0874	0.0903	-0.0116
benefit_cosmetics	0.0781	0.0875	-0.0701	0.0781	0.0807	-0.0200
morphe	0.0803	0.0923	-0.0548	0.0803	0.0813	-0.0050
loreal	0.0726	0.0850	-0.0935	0.0726	0.0745	-0.0141
nars	0.0684	0.0752	-0.0534	0.0684	0.0703	-0.0143

colourpop_cosmetics	0.0565	0.0700	-0.1102	0.0565	0.0575	-0.0080
becca_cosmetics	0.0478	0.0589	-0.1056	0.0478	0.0482	-0.0043
laura_mercier	0.0515	0.0575	-0.0556	0.0515	0.0528	-0.0117
makeup_forever	0.0517	0.0620	-0.0914	0.0517	0.0524	-0.0066
kat_von_d	0.0413	0.0542	-0.1377	0.0413	0.0420	-0.0071
it_cosmetics	0.0496	0.0547	-0.0468	0.0496	0.0490	0.0056
smashbox	0.0444	0.0500	-0.0583	0.0444	0.0449	-0.0047
elf_cosmetics	0.0364	0.0450	-0.0815	0.0364	0.0365	-0.0004
wet_n_wild	0.0286	0.0412	-0.1593	0.0286	0.0292	-0.0084
hourglass	0.0348	0.0386	-0.0448	0.0348	0.0351	-0.0042
milani	0.0244	0.0359	-0.1759	0.0244	0.0251	-0.0100
marc_jacobs	0.0334	0.0364	-0.0358	0.0334	0.0342	-0.0093
rimmel_london	0.0261	0.0324	-0.0867	0.0261	0.0273	-0.0163
covergirl	0.0272	0.0323	-0.0616	0.0272	0.0278	-0.0070
la_girl	0.0292	0.0287	0.0048	0.0292	0.0305	-0.0120
beauty_blender	0.0300	0.0297	0.0030	0.0300	0.0298	0.0017
huda_beauty	0.0246	0.0270	-0.0302	0.0246	0.0249	-0.0033
jouer	0.0310	0.0332	-0.0173	0.0310	0.0325	-0.0115
eye_other	0.3038	0.3449	-0.1368	0.3038	0.3084	-0.0152
foundation	0.2552	0.2864	-0.1142	0.2552	0.2580	-0.0104
lash	0.2495	0.2853	-0.1173	0.2495	0.2526	-0.0104
brow	0.2374	0.2626	-0.0976	0.2374	0.2410	-0.0138
brush	0.2487	0.2748	-0.0776	0.2487	0.2492	-0.0014
palette	0.2325	0.2773	-0.1791	0.2325	0.2355	-0.0119
powder	0.2138	0.2391	-0.1099	0.2138	0.2173	-0.0155
conceal	0.1972	0.2253	-0.1280	0.1972	0.2007	-0.0161
tool	0.1828	0.2125	-0.1299	0.1828	0.1847	-0.0083
liner	0.1839	0.2173	-0.1539	0.1839	0.1868	-0.0130
lip_stick	0.1793	0.2079	-0.1327	0.1793	0.1829	-0.0168
shadow	0.1657	0.2003	-0.1580	0.1657	0.1679	-0.0100
blush	0.1630	0.1953	-0.1620	0.1630	0.1660	-0.0153
highlight	0.1685	0.1886	-0.0922	0.1685	0.1722	-0.0169
mascara	0.1613	0.1851	-0.1204	0.1613	0.1647	-0.0173
bronz	0.1400	0.1586	-0.0981	0.1400	0.1427	-0.0138
primer	0.1326	0.1524	-0.1152	0.1326	0.1349	-0.0130
lip_other	0.1138	0.1299	-0.1054	0.1138	0.1152	-0.0093
contour	0.1115	0.1322	-0.1113	0.1115	0.1140	-0.0135
oil	0.1155	0.1096	0.0343	0.1155	0.1141	0.0082
spray	0.0908	0.0906	0.0011	0.0908	0.0889	0.0131
mask	0.0615	0.0637	-0.0181	0.0615	0.0621	-0.0052

moisturi	0.0457	0.0457	-0.0009	0.0457	0.0463	-0.0075
serum	0.0421	0.0428	-0.0064	0.0421	0.0431	-0.0093
cleans	0.0452	0.0455	-0.0026	0.0452	0.0456	-0.0041
lip_gloss	0.0397	0.0422	-0.0260	0.0397	0.0415	-0.0190
mist	0.0401	0.0439	-0.0379	0.0401	0.0420	-0.0186
comb	0.0370	0.0286	0.0650	0.0370	0.0344	0.0205
fac_other	0.0254	0.0288	-0.0441	0.0254	0.0258	-0.0056
spong	0.0316	0.0336	-0.0207	0.0316	0.0319	-0.0032
Monthly ad. Spending	6.2974	6.2076	0.0408	6.2974	6.3162	-0.0085
Notes: The variable distance is the propensity score.						

OA 3.B. MEAN INFLUENCER VARIABLES BEFORE AND AFTER PSM

To evaluate the effectiveness of the propensity score matching (PSM), we assess the statistical significance of content variables in the matched and unmatched samples using t-tests, following the method proposed by Montaguti et al. (2015). PSM is deemed successful if the t-statistics for the matched sample are not statistically significant, indicating that the matching procedure has effectively reduced observable differences between the treated and control groups.

Table S22 presents the means and t-statistics of 153 out of a total of 1,364 variables used in the propensity score matching (PSM) analysis. For simplicity, we have not included the influencer and time fixed effects. Our PSM procedure employs an extensive set of 1,364 matching variables, and we find no significant differences between the treated and control groups in the matched sample, except for one verbal topic dummy variable among ten dummies. Based on this finding, we conclude that the PSM analysis is successful in eliminating observable differences between the treated and control groups.

Table S22. Mean & t-Statistics of Content Variables in the All & Matched Data

Variable	(A) All Means -Treated	(B) All Means - Control	(C) Matched Means - Treated	(D) Matched Means - Control	(A) vs (B) tstat	(C) vs. (D) tstat
N. of videos	2.8	2.843	2.8	2.802	-0.348	-7.691***
N. of sponsored videos	1.247	0.844	1.247	1.236	1.409	63.616***
N. of sponsored fashion	0.314	0.187	0.314	0.319	-0.952	30.273***
N. of sponsored hair	0.235	0.12	0.235	0.224	2.09*	29.283***
N. of sponsored makeup	0.572	0.422	0.572	0.573	-0.173	29.345***
N. of sponsored other	0.435	0.275	0.435	0.434	0.107	32.569***
N. of sponsored skincare	0.159	0.115	0.159	0.161	-0.419	16.131***
Duration (Seconds)	6.444	6.426	6.444	6.447	-0.522	3.735***
View count	10.848	10.761	10.848	10.857	-0.56	6.861***
Like count	7.486	7.419	7.486	7.5	-0.936	5.876***
Dislike count	3.877	3.817	3.877	3.89	-0.944	5.807***

Comment count	5.075	5.096	5.075	5.092	-1.319	-2.048*
Title length (N. of character)	3.925	3.898	3.925	3.923	0.626	9.252***
Description length (N. of character)	7.348	7.287	7.348	7.349	-0.022	10.245***
Tenure (Number of days)	7.07	6.851	7.07	7.087	-1.844	29.154***
Follower count (Raw)	12.122	11.77	12.122	12.141	-1.22	27.861***
Total views (Raw)	18.023	17.779	18.023	18.011	0.706	18.755***
IG follower count (Raw)	10.735	10.555	10.735	10.762	-1.199	9.967***
Follower count (Difference)	-0.016	0.003	-0.016	-0.006	-0.918	-2.385*
Total views (Difference)	-0.07	0.013	-0.07	-0.059	-1.045	-10.193***
IG follower count (Diff.)	0.015	-0.003	0.015	0.017	-0.153	2.235*
Format - GRWM (dummy)	0.171	0.146	0.171	0.173	-0.713	12.091***
Format - haul (dummy)	0.377	0.354	0.377	0.378	-0.485	7.94***
Format - review (dummy)	0.317	0.348	0.317	0.316	0.19	-10.871***
Format - routine (dummy)	0.263	0.255	0.263	0.266	-0.961	3.137**
Format - tutorial (dummy)	0.692	0.702	0.692	0.692	0.008	-3.822***
Format - vlog (dummy)	0.413	0.389	0.413	0.412	0.102	7.048***
Spoken unigram frequency	6.981	6.857	6.981	6.98	0.064	10.702***
Spoken bigram frequency	5.277	5.19	5.277	5.278	-0.104	9.892***
Spoken trigram frequency	3.109	3.061	3.109	3.11	-0.128	9.03***
Verbal concreteness	1.311	1.288	1.311	1.311	0.035	10.825***
Verbal topic 0	0.083	0.06	0.083	0.076	3.713***	16.387***
Verbal topic 1	0.012	0.015	0.012	0.012	0.149	-9.065***
Verbal topic 2	0.008	0.013	0.008	0.009	-0.981	-14.019***
Verbal topic 3	0.013	0.01	0.013	0.013	-0.253	8.626***
Verbal topic 4	0.009	0.009	0.009	0.009	-0.263	0.733
Verbal topic 5	0.008	0.008	0.008	0.008	-0.609	0.155
Verbal topic 6	0.006	0.008	0.006	0.007	-0.741	-8.968***
Verbal topic 7	0.007	0.008	0.007	0.007	-0.38	-4.673***
Verbal topic 8	0.005	0.006	0.005	0.005	-0.325	-6.665***
Verbal topic 9	0.004	0.005	0.004	0.005	-0.221	-3.948***
Age	3.436	3.419	3.436	3.436	0.09	8.406***
Angry (%)	0.059	0.061	0.059	0.058	0.436	-5.161***
Disgust (%)	0.013	0.013	0.013	0.012	0.525	-1.015
Fear (%)	0.101	0.105	0.101	0.101	0.305	-4.842***
Happy (%)	0.28	0.263	0.28	0.28	-0.129	11.531***
Neutral (%)	0.33	0.326	0.33	0.33	-0.133	2.758**
Sad (%)	0.164	0.173	0.164	0.165	-0.419	-8.199***
Surprise (%)	0.049	0.05	0.049	0.049	0.415	-1.456
Asian (%)	0.154	0.143	0.154	0.154	0.122	6.018***

Black (%)	0.164	0.123	0.164	0.161	0.966	16.678***
Indian (%)	0.059	0.058	0.059	0.061	-1.318	2.327*
Latino/Hispanic (%)	0.157	0.158	0.157	0.158	-1.463	-0.971
Middle Eastern (%)	0.126	0.134	0.126	0.127	-0.664	-8.99***
White (%)	0.336	0.375	0.336	0.336	0.005	-16.589***
Facial attractiveness	1.358	1.355	1.358	1.358	-0.346	3.035**
Makeup heaviness	0.045	0.041	0.045	0.045	0.703	15.674***
Top, t-shirt, sweatshirt	0.013	0.012	0.013	0.013	-0.038	3.648***
Jacket	0	0	0	0	0.836	-1.147
Pants	0.003	0.003	0.003	0.003	-0.302	0.258
Skirt	0	0	0	0	-0.855	1.346
Dress	0.134	0.133	0.134	0.134	0.256	0.55
Glasses	0.003	0.004	0.003	0.004	-0.563	-0.97
Hat	0	0	0	0	-0.656	1.548
Watch	0	0	0	0	-0.038	-2.446*
Belt	0	0	0	0	0.401	0.225
Tights, stockings	0	0	0	0	-1.575	-2.635**
Shoe	0.068	0.067	0.068	0.068	-0.84	1.071
Bag, wallet	0	0	0	0	1.278	2.203*
Collar	0	0	0	0	-0.588	-0.03
Lapel	0	0	0	0	-0.307	-0.143
Sleeve	0.525	0.515	0.525	0.525	-0.013	4.261***
Pocket	0.001	0.001	0.001	0.001	-0.472	0.716
Neckline	0.19	0.181	0.19	0.19	-0.312	6.688***
Foundation1 (G)	0.429	0.427	0.429	0.43	-0.274	1.577
Foundation2 (G)	0.414	0.412	0.414	0.415	-0.158	1.568
Foundation3 (G)	0.415	0.412	0.415	0.415	-0.046	1.669
Blush1 (G)	0.36	0.361	0.36	0.361	-0.145	-0.777
Blush2 (G)	0.363	0.364	0.363	0.363	-0.141	-0.791
Blush3 (G)	0.384	0.384	0.384	0.384	0.007	0.337
Blush4 (G)	0.407	0.405	0.407	0.407	-0.005	1.306
Blush5 (G)	0.424	0.421	0.424	0.424	-0.008	1.603
Lip (G)	0.304	0.294	0.304	0.304	0.324	8.653***
Lipliner (G)	0.338	0.333	0.338	0.338	0.208	4.359***
Eyeshadow1 (G)	0.344	0.34	0.344	0.344	-0.217	2.281*
Eyeshadow2 (G)	0.355	0.352	0.355	0.356	-0.161	2.547*
Eyeshadow3 (G)	0.334	0.33	0.334	0.334	-0.06	3.357***
Eyeshadow4 (G)	0.317	0.312	0.317	0.317	0.046	4.054***
Eyeshadow5 (G)	0.305	0.3	0.305	0.305	-0.004	4.157***
Eyeliner (G)	0.112	0.11	0.112	0.112	0.262	3.783***

Eyebrow (G)	0.299	0.299	0.299	0.3	-0.299	-0.045
NIMA average score	1.807	1.779	1.807	1.809	-0.445	10.281***
mac cosmetics	0.132	0.147	0.132	0.133	-0.58	-8.991***
anastasia beverly hills	0.091	0.117	0.091	0.094	-1.411	-19.391***
nyx cosmetics	0.083	0.101	0.083	0.084	-0.834	-14.137***
maybelline	0.083	0.101	0.083	0.086	-1.814	-14.485***
too faced	0.092	0.102	0.092	0.093	-1.004	-7.67***
urban decay	0.086	0.103	0.086	0.088	-0.904	-14.317***
tarte	0.088	0.1	0.088	0.089	-0.468	-8.959***
ofra	0.087	0.096	0.087	0.09	-1.059	-3.873***
benefit cosmetics	0.078	0.087	0.078	0.081	-1.826	-8.264***
morphe	0.08	0.092	0.08	0.081	-0.456	-6.527***
loreal	0.073	0.085	0.073	0.075	-1.287	-11.016***
nars	0.068	0.075	0.068	0.07	-1.312	-6.314***
colourpop cosmetics	0.057	0.07	0.057	0.057	-0.729	-12.805***
becca cosmetics	0.048	0.059	0.048	0.048	-0.392	-12.442***
laura mercier	0.051	0.058	0.051	0.053	-1.071	-6.539***
makeup forever	0.052	0.062	0.052	0.052	-0.61	-10.748***
kat von d	0.041	0.054	0.041	0.042	-0.645	-15.939***
it cosmetics	0.05	0.055	0.05	0.049	0.519	-5.519***
smashbox	0.044	0.05	0.044	0.045	-0.437	-6.904***
elf cosmetics	0.036	0.045	0.036	0.036	-0.039	-9.63***
wet n wild	0.029	0.041	0.029	0.029	-0.761	-18.115***
hourglass	0.035	0.039	0.035	0.035	-0.386	-5.3***
milani	0.024	0.036	0.024	0.025	-0.903	-19.853***
marc jacobs	0.033	0.036	0.033	0.034	-0.859	-4.269***
rimmel london	0.026	0.032	0.026	0.027	-1.46	-10.174***
covergirl	0.027	0.032	0.027	0.028	-0.643	-7.268***
la girl	0.029	0.029	0.029	0.03	-1.083	0.582
beauty blender	0.03	0.03	0.03	0.03	0.158	0.356
huda beauty	0.025	0.027	0.025	0.025	-0.308	-3.604***
jouer	0.031	0.033	0.031	0.032	-1.031	-2.063*
eye other	0.304	0.345	0.304	0.308	-1.396	-16.217***
foundation	0.255	0.286	0.255	0.258	-0.957	-13.551***
lash	0.249	0.285	0.249	0.253	-0.951	-13.918***
brow	0.237	0.263	0.237	0.241	-1.27	-11.565***
brush	0.249	0.275	0.249	0.249	-0.128	-9.233***
palette	0.233	0.277	0.233	0.236	-1.093	-21.105***
powder	0.214	0.239	0.214	0.217	-1.421	-13.002***
conceal	0.197	0.225	0.197	0.201	-1.474	-15.123***

tool	0.183	0.212	0.183	0.185	-0.763	-15.341***
liner	0.184	0.217	0.184	0.187	-1.193	-18.087***
lip stick	0.179	0.208	0.179	0.183	-1.532	-15.689***
shadow	0.166	0.2	0.166	0.168	-0.921	-18.589***
blush	0.163	0.195	0.163	0.166	-1.397	-18.933***
highlight	0.168	0.189	0.168	0.172	-1.537	-10.956***
mascara	0.161	0.185	0.161	0.165	-1.582	-14.175***
bronz	0.14	0.159	0.14	0.143	-1.26	-11.567***
primer	0.133	0.152	0.133	0.135	-1.186	-13.571***
lip other	0.114	0.13	0.114	0.115	-0.855	-12.433***
contour	0.112	0.132	0.112	0.114	-1.216	-13.103***
oil	0.115	0.11	0.115	0.114	0.761	4.114***
spray	0.091	0.091	0.091	0.089	1.235	0.135
mask	0.061	0.064	0.061	0.062	-0.472	-2.128*
moisturi	0.046	0.046	0.046	0.046	-0.683	-0.104
serum	0.042	0.043	0.042	0.043	-0.853	-0.765
cleans	0.045	0.045	0.045	0.046	-0.381	-0.306
lip gloss	0.04	0.042	0.04	0.042	-1.71	-3.128**
mist	0.04	0.044	0.04	0.042	-1.692	-4.519***
comb	0.037	0.029	0.037	0.034	1.973*	7.984***
fac other	0.025	0.029	0.025	0.026	-0.512	-5.122***
spong	0.032	0.034	0.032	0.032	-0.285	-2.486*
Monthly ad. Spending	6.297	6.208	6.297	6.316	-0.793	4.837***

OA 3.C. DO INFLUENCERS ADJUST VERBAL AND VISUAL FEATURES IN SPONSORED VIDEOS?

We apply Equation (2) from the main paper to investigate the visual and verbal features of sponsored and matched non-sponsored video data. The estimation is repeated with fixed effects for the influencer, time, visual, verbal, text, popularity, and video format, excluding the variable used as the dependent variable. Table S23 presents the findings, indicating no significant differences between sponsored and non-sponsored videos on visual or verbal variables in the sponsored and matched non-sponsored data.

Table S23. Change in Verbal/Visual Features in Videos in the Treated Influencer-Time Panels

		Matched		
Group	Dependent variable	Coef	SE	Pval
Verbal	sophistication - 1gram	0.00007	0.00118	0.953
	sophistication - 2gram	0.00112	0.00079	0.1557
	sophistication - 3gram	-0.00084	0.00102	0.4122
	sophistication - concreteness	-0.00028	0.0002	0.1752

	verbal_topic_0	0.00049	0.00083	0.5557
	verbal_topic_1	0.0003	0.00033	0.3536
	verbal_topic_2	-0.00034	0.00032	0.288
	verbal_topic_3	-0.00003	0.00034	0.9258
	verbal_topic_4	0.00025	0.00029	0.3924
	verbal_topic_5	0	0.00034	0.9981
	verbal_topic_6	0.00007	0.00023	0.7484
	verbal_topic_7	-0.00001	0.00026	0.9675
	verbal_topic_8	0.00005	0.0002	0.7995
	verbal_topic_9	-0.00013	0.00023	0.5612
Visual	age	0.00004	0.00034	0.9141
	emotion_angry	0	0	0.8224
	emotion_disgust	0	0	0.8224
	emotion_fear	0	0	0.8224
	emotion_happy	0	0	0.8224
	emotion_neutral	0	0	0.8224
	emotion_sad	0	0	0.8224
	emotion_surprise	0	0	0.8224
	race_asian	0	0	0.8224
	race_black	0	0	0.8224
	race_indian	0	0	0.8224
	race_latino_hispanic	0	0	0.8224
	race_middle_eastern	0	0	0.8224
	race_middle_white	0	0	0.8224
	facial attractiveness	-0.0004	0.00034	0.24266
	makeup heaviness	-0.00003	0.00009	0.76567
	object_top_t-shirt_sweatshirt	0.00022	0.00037	0.54954
	object_jacket	-0.00002	0.00001	0.23172
	object_pants	0.00015	0.00017	0.37883
	object_skirt	-0.00001	0.00002	0.43247
	object_dress	-0.00034	0.00115	0.76547
	object_glasses	-0.00002	0.00018	0.90418
	object_hat	-0.00004	0.00003	0.18504
	object_watch	0	0	0.42095
	object_belt	0.00002	0.00001	0.05561
	object_tights_stockings	-0.00003	0.00006	0.65944
	object_shoe	-0.00023	0.00085	0.78491
	object_bag_wallet	-0.00002	0.00006	0.76121
	object_collar	0.00004	0.00006	0.48656
	object_lapel	0.00003	0.00006	0.66283

	object_sleeve	0.00045	0.00145	0.75454
	object_pocket	-0.00007	0.00009	0.46197
	object_neckline	0.00108	0.00124	0.38564
	foundation1 (G)	0.00009	0.00025	0.72003
	foundation2 (G)	-0.00012	0.00022	0.58363
	foundation3 (G)	0.00008	0.00026	0.74758
	blush1 (G)	-0.00002	0.00015	0.87317
	blush2 (G)	-0.00006	0.0001	0.5729
	blush3 (G)	0.00003	0.00007	0.61432
	blush4 (G)	-0.00004	0.00005	0.43751
	blush5 (G)	0.0001	0.00008	0.23746
	lip (G)	0.00011	0.0003	0.70391
	lipliner (G)	-0.00006	0.00019	0.73016
	eyeshadow1 (G)	-0.00007	0.00027	0.78607
	eyeshadow2 (G)	0.00006	0.00011	0.57939
	eyeshadow3 (G)	-0.00003	0.00007	0.67983
	eyeshadow4 (G)	0.00001	0.00006	0.91981
	eyeshadow5 (G)	0	0.00009	0.96016
	eyeliner (G)	0.00013	0.00028	0.63922
	eyebrow (G)	-0.00024	0.00032	0.44741
	NIMA average score	0.0014	0.0018	0.43537

OA 3.D. PSM: TWO SENSITIVITY ANALYSES

Rosenbaum bounds analysis for sensitivity to unobservables

Incorporating an exhaustive list of content variables is a common strategy to control for potential confounding factors in observational studies. However, such variables may not capture all the relevant information, leading to a potential bias in the estimated effects. To address this issue, we adopt the bounding approach suggested by Rosenbaum (2002) to quantify the magnitude of the omitted variable bias. Specifically, we use non-experimental data to establish a plausible range of unobservable factors that might confound the effect of sponsorship on the outcome of interest. By comparing the estimated treatment effect with the lower and upper bounds of the unobserved confounding, we can assess the robustness of our results to potential omitted variable bias.

Rosenbaum's bounding approach is a useful tool to assess the potential bias caused by unobservable characteristics that are not captured by the included content variables in the matching procedure. This approach evaluates the sensitivity of the Conditional Independence Assumption (CIA) to omitted variables, which assumes that the treatment and the outcome are independent conditional on the propensity score and covariates (Caliendo and Kopeinig 2008). Thus, if there is a presence of omitted variable bias, two influencers with the same propensity score may have different probabilities of receiving treatment. To test the violation of the CIA, we calculate the odds ratio of treatment between two influencers (i.e., Influencer 1 and Influencer 2) who have the same observed content variables but may differ on unobserved characteristics that could influence the likelihood of treatment. Caliendo and Kopeinig (2008) suggest the following odds ratio formula:

$$\frac{odds\ of\ 1}{odds\ of\ 2} = \frac{\frac{pscore_1}{(1 - pscore_1)}}{\frac{pscore_2}{(1 - pscore_2)}} = \frac{\exp(\beta X_1 + \gamma U_1)}{\exp(\beta X_2 + \gamma U_2)} = \frac{\exp(\beta X_1 + \gamma U_1)}{\exp(\beta X_1 + \gamma U_2)} = \exp(\gamma(U_1 - U_2))$$

where *pscore* is the propensity score, *X* is the set of content variables (with weight β), and *U* is the set of unobserved characteristics (with weight γ). The odds ratio deviates from 1 if the influencers have different unobserved characteristics (U_1, U_2). The Rosenbaum bounds for the effect of sponsorship on consumer sentiment are presented in Table S24.

Table S24. Rosenbaum Bounds of the Effect of Sponsorship on Consumer Sentiment

gamma	sig+	sig-	t-hat+	t-hat-	CI+	CI-
1	0	0	-.022813	-.022813	-.025946	-.019681
1.05	0	0	-.027214	-.018412	-.030351	-.015277
1.1	0	0	-.031412	-.014219	-.034545	-.011077
1.15	0	1.1e-10	-.035415	-.010205	-.038556	-.007064
1.2	0	.000038	-.039249	-.006371	-.042399	-.003221
1.25	0	.047145	-.04293	-.002692	-.046085	.000461
1.3	0	**.699438**	-.046463	.000839	-.049621	.003998
1.35	0	.995806	-.049857	.004235	-.053016	.007406
1.4	0	.999999	-.053119	.00751	-.056287	.010682
1.45	0	1	-.056266	.010662	-.059442	.013842
1.5	0	1	-.059304	.013703	-.062485	.016894
1.55	0	1	-.062235	.016643	-.065426	.019844
1.6	0	1	-.065073	.01949	-.06827	.022698

* gamma - log odds of differential assignment due to unobserved factors
sig+ - upper bound significance level
sig- - lower bound significance level
t-hat+ - upper bound Hodges-Lehmann point estimate
t-hat- - lower bound Hodges-Lehmann point estimate

Our estimated treatment effect is negative, and we are particularly interested in the lower bound of the confidence interval (given by t-hat-). The confidence interval includes 0 at a gamma value of 1.3, indicating that unobserved covariates may affect the odds of treatment by up to 1.3 times and negate our treatment effect. However, our gamma value is higher than the lowest allowed log odds of 1.15 in the examples of DiPrete and Gangl (2004), indicating that our estimation of the treatment effect on the PSM sample is satisfactorily robust to omitted variables. Higher values of gamma indicate a more robust treatment effect.

OA 3.E. INSTRUMENTAL VARIABLE VALIDITY CHECK

Instrument relevance test

To ensure the validity of instrumental variable estimation, it is essential to meet the relevance condition. This condition requires that the instrumental variable is strongly correlated with the endogenous regressor (Conley et al. 2012; Rossi 2014). To test the relevance of our instrument, we employ the F-statistics test by comparing the proposed model (with both exogenous regressors and instruments) with the null model (with only exogenous regressors). If the F-statistics test produces a significant result, our instrument has high relevance to the regressor, indicating that we have a strong instrument. In our study, we utilize sponsorship as the dependent variable for both the proposed model in Equation (S9) and the null model in Equation (S10).

$$\begin{aligned} Sponsorship_{ijt} &= \beta_0 + \overrightarrow{\beta_1} \times \overrightarrow{Concurrent\ videos\ (m, m')} \\ &+ \overrightarrow{\beta_2} \times \overrightarrow{Voice_{ijt}^{k}} \times \overrightarrow{Concurrent\ videos\ (m, m')} \\ &+ \overrightarrow{\Gamma} \times \overrightarrow{ControlVariables} + I_i + \tau_t + \epsilon_{ijt} \end{aligned} \quad \text{(S9)}$$

$$Sponsorship_{ijt} = \beta_0 + \overrightarrow{\beta_1} \times \overrightarrow{Voice_{ijt}^{k}} + \overrightarrow{\Gamma} \times \overrightarrow{Control\ variables} + I_i + \tau_t + \epsilon_{ijt} \quad \text{(S10)}$$

Table S25 shows the results of the instrumental variable test for the relevance of the instruments (the number of sponsoring brand and the parent company) to the sponsorship treatment. The F-statistics of the proposed model and the null model are computed and compared to determine the strength of the instruments. A significant F-statistic indicates a strong correlation between the instruments and the endogenous regressors, suggesting that the instruments meet the relevance condition. We observe that the F-statistics are significant at a 0.01 level for all six endogenous regressors, indicating that the selected instrument is highly relevant to the sponsorship treatment.

Table S25. Instrument Relevance Statistics of the Six Endogenous Regressors

Variable	F-statistics	P-value
Sponsorship	1,311.377	2.2e-16***
*Sponsorship * Loudness*	1,730.279	2.2e-16***
*Sponsorship * Pitch*	1,484.795	2.2e-16***
*Sponsorship * Loudness variability*	1,500.133	2.2e-16***
*Sponsorship * Pitch variability*	1,480.335	2.2e-16***
*Sponsorship * Talking duration*	1,498.509	2.2e-16***

Hausman-Wu test for exogeneity

The use of the IV approach is motivated by endogeneity concerns associated with sponsorship. However, if sponsorship is exogenous, the use of an instrument is not necessary. In the case of an exogenous regressor, introducing an instrument would result in inefficiencies in the IV estimates, and both OLS and IV estimates would be consistent, as noted by Green (2003). However, if the regressor is indeed endogenous, the IV estimates would remain consistent while the OLS estimates would not. Hence, the use of an instrument in this scenario is critical to achieving reliable estimates.

Formally, the Hausman-Wu test examines the following two hypotheses:

H0) Both the OLS and IV estimates are consistent, and the OLS estimate is more efficient.
H1) The IV estimate is consistent, while the OLS estimate is inconsistent.

We use the following model specification to test the hypotheses:

$$\begin{aligned} Sentiments_{ijt} \\ &= \beta_0 + \beta_1 \times Sponsorship_{ijt} + \beta_2 \times \overrightarrow{Voice_{ijt}^{k}} \\ &+ \beta_3 \times Sponsorship_{ijt} \times \overrightarrow{Voice_{ijt}^{k}} + \widetilde{Sponsorship_{ijt}} \\ &+ \sum_{k} \widetilde{Sponsored\ voice_{ijt}} + \vec{\Gamma} \times \overrightarrow{Control\ variables} + \epsilon_{ijt} \end{aligned} \tag{S11}$$

$$Sentiments_{ijt} = \beta_0 + \vec{\Gamma} \times \overrightarrow{Control\ variables} + \epsilon_{ijt} \tag{S12}$$

In Equation (11), $\widetilde{Sponsorship_{ijt}}$ and $\sum_{j} \widetilde{Sponsored\ voice_{ijt}}$ represent the residuals of the first-stage regression, with sponsorship and the interactions between sponsorship and the vocal characteristics as the dependent variables. The endogeneity of sponsorship raises concerns regarding the accuracy of the OLS estimates, thus necessitating the use of instruments. The Hausman-Wu test statistic, which measures the difference between the OLS and IV estimates, is 2.573, and the p-value is 0.0171. As Hausman-Wu test statistics are typically significant at $p < 0.05$, we reject the null hypothesis, indicating that OLS estimates are inconsistent without instruments, and sponsorship is indeed endogenous.

Sargan test for over-identifying restrictions

To verify the validity of our instruments, we employ the Sargan test. This test assesses whether the residual obtained from the second-stage regression is correlated with the regressors, indicating over-identification. Specifically, the Sargan test statistic is calculated as N multiplied by the R-squared value of the regression where the dependent variable is the residual, with N being the total number of observations (32,920) and R^2 equaling 1.74058 x 10-4.

The Sargan test statistic follows an asymptotic chi-squared distribution, and we find that it equals 5.730 with a p-value of 0.4541. Given that this p-value is higher than the threshold of 0.05, we fail to reject the null hypothesis that all of our instruments are valid. We thus conclude that our instruments are indeed valid for our analysis.

OA SECTION 4. FUZZY REGRESSION DISCONTINUITY APPROACH

OA 4.A. SUPPORT FOR THE ASSIGNMENT VARIABLES IN THE RD MODEL

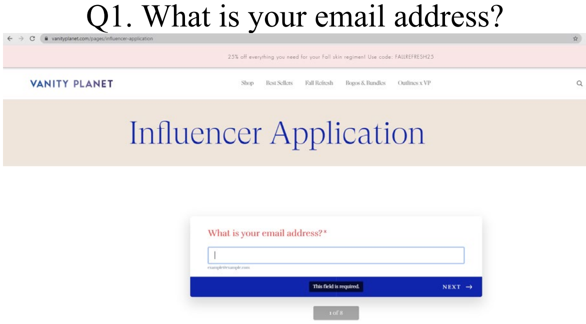
As indicated in the manuscript, we focused on the ten cosmetics brands with the highest number of sponsorships in our dataset, including Fabfitfun, Garnier, L'Oreal Paris, Maybelline, Neutrogena, Olay, Scentbird, Sephora, Ulta, and Vanity Planet. We verified that these brands use two common criteria to select influencers, namely their follower count and average number of views. In this section, we present additional evidence to support our findings, including influencer applications, email correspondence with a marketing manager, and interviews with company executives.

1. Influencer Applications

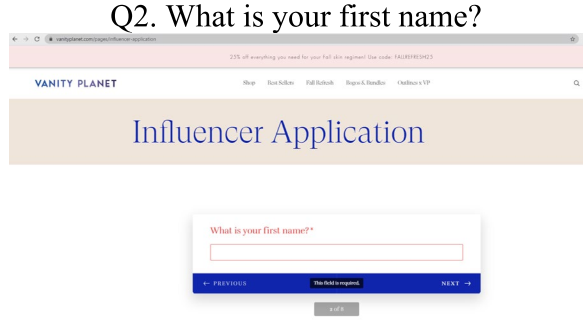
   1) Vanity Planet

   The influencer application of Vanity Planet includes nine questions, among which questions four and six pertain to the assignment variables under investigation, i.e., the number of followers (Q4) and the number of views (Q6). The remaining seven questions relate to non-performance-related information such as email and social media address. These findings indicate that Vanity Planet relies on the assignment variables to select influencers for sponsorship.

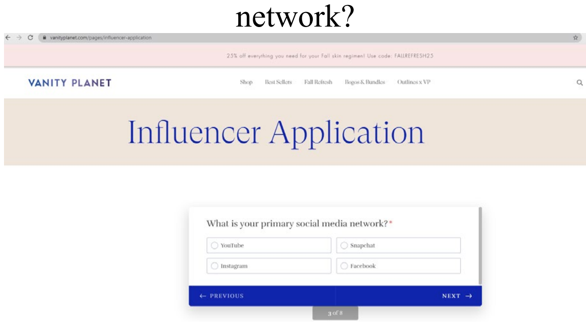
Q1. What is your email address?



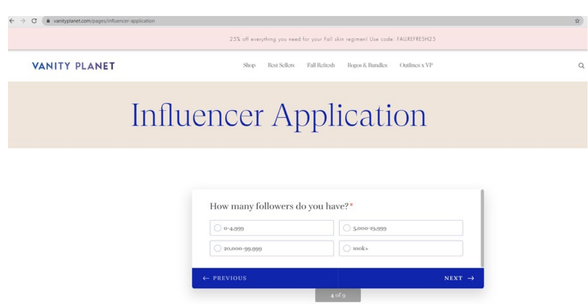
Q2. What is your first name?



Q3. What is your primary social media network?

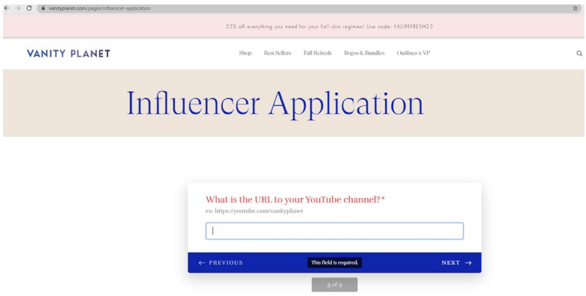


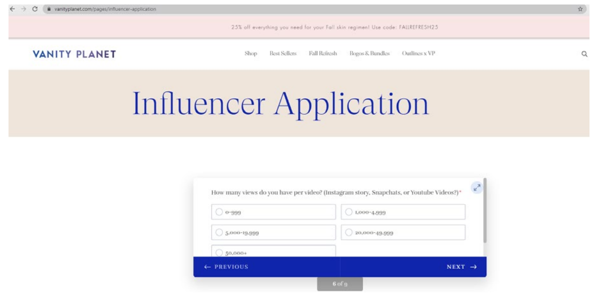
Q4. How many ***followers*** do you have?



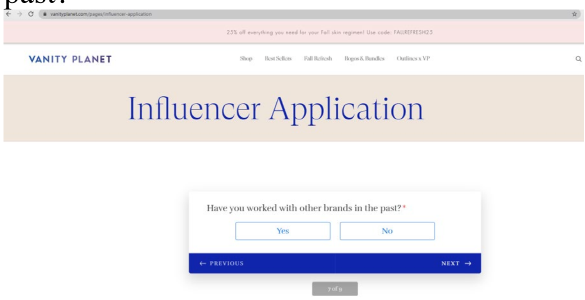
Q5. What is the URL to your YouTube/Snapchat/Instagram/Facebook channel?

Q6. How many ***views*** do you have per video? (Instagram story, Snapchats, or Youtube videos?)

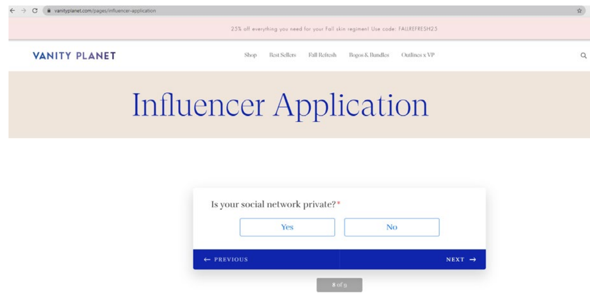




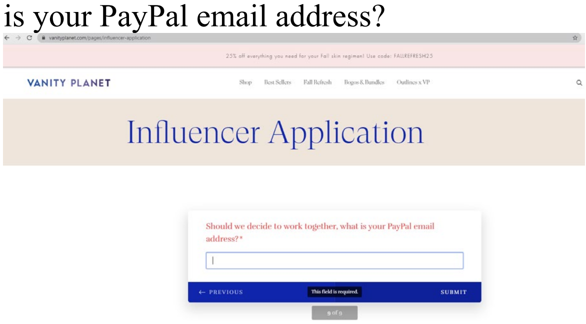
Q7. Have you worked with other brands in the past?



Q8. Is your social network private?



Q9. Should we decide to work together, what is your PayPal email address?



2) Sephora & Ulta

Following a similar practice to Vanity Planet, Sephora and Ulta require influencer applicants to provide information on their follower count and average views. Specifically, Sephora solicits this information through their registration page on the Linkshare platform (https://signup.linkshare.com/publishers/registration/landing?ls-locale=us&host=linkshare&mid=2417), while Ulta's application process asks for such information through their campaign mediapartner sign-up page on the Impact platform (https://app.impact.com/campaign-mediapartner-signup/Ulta.brand?execution=e2s1). This evidence supports the notion that these brands also employ our assignment variables as a basis for selecting influencers for sponsorship.

Email response from the Neutrogena PR/marketing manager

In September 2020, we reached out to Michelle, the PR/marketing manager at Johnson & Johnsons Customer Care center, to inquire about the influencer selection process for brands under Johnson & Johnsons, including Neutrogena and Aveeno. Michelle disclosed that the brands seek out influencers whose follower count exceeds a brand-specific threshold, which aligns with one of the assignment variables.

Johnson & Johnson Consumer Inc. Reference - 002050731

Neutrogena Consumer Care <consumer-us-ntg@its.jnj.com> Sat, Sep 5, 7:02 AM
to me

Dear ,

Thank you for taking the time to contact us at NEUTROGENA®.

In order to be considered by J&J/Brand the influencer/blogger/celebrity must meet the following criteria:
5000+ followers on at least one social platform
For US Pain Brands only consider Instagram followers
Reside in the same country as the brand they want to work with.

While we do not have the exact numbers, these criteria are what we are looking for a person who wants to partner with or be sponsored by a J&J brand, wants to review a J&J product or sample on their blog or on any social media platform, requests products/samples to review on their blog and/or social media platform.

Serim, we appreciate you reaching out to us. If you have any comments or questions about our products in the future, please contact us through our website www.neutrogena.com or by phone 1-800-582-4048. Our specialists are available Monday through Friday, 9:00 a.m. to 5:30 p.m., ET and will be happy to help you.

Sincerely,

Michelle
Consumer Care Center
002050731

ref:_00D58ZcdK._5001iSfHq0:ref

2. Articles and Interviews with Executives

1) Scentbird

Sergei Gusev, the co-founder and Chief Operating Officer at Scentbird, has discussed Scentbird's influencer marketing approach in two articles: "How we got from $0 to $75,000 MRR with zero marketing budget" (https://medium.com/@ligeo/how-we-got-from-0-to-75-000-mrr-with-zero-marketing-budget-b20101b09a76) and "How Influencer Marketing helped us grow from $0 to $700k+ monthly revenue" (https://medium.com/@ligeo/how-influencer-marketing-helped-us-grew-from-0-to-700k-monthly-revenue-51644e79f7a9). In these articles, Gusev revealed that Scentbird selects influencers based on a brand-specific threshold of the number of followers, which is consistent with one of the assignment variables used in this study.

2) Fabfitfun

During the 2019 Digital Marketing World Forum (DMWF), Katie Gagnon, FabFitFun's manager of influencer marketing and talent partnerships, presented on the process of influencer marketing campaigns (https://www.warc.com/newsandopinion/news/how-fabfitfun-taps-influencer-power/41651). Gagnon mentioned, "We aim for about 500 new micro influencers per season and hope to have around 500 to 750 go live each season. So many people are influencers at this point, so we've definitely got a little bit tighter in terms of follower count." This implies that Fabfitfun selects influencers for sponsorship based on their follower count.

3) L'Oreal Group Brands: L'Oreal Paris, Maybelline, Garnier

According to a PRweek article (https://www.prweek.com/article/1668404/why-micro-influencers-drive-cross-border-success), three brands under the L'Oreal group, namely L'Oreal Paris, Maybelline, and Garnier, have been observed to collaborate with influencers who meet specific follower and view count thresholds.

4) Olay

According to a report from Tribe Group, a third-party influencer marketing platform for brands and agencies (https://www.tribegroup.co/), Olay selects influencers based on two key criteria - the reach of the influencer (i.e., the number of followers they have) and their engagement rate (i.e., the average views of their content). This finding provides further evidence of the importance of follower count and views as primary assignment variables for influencer sponsorships across a range of cosmetics brands.

OA 4.B. FUZZY REGRESSION DISCONTINUITY (FRD) PROCEDURE

For each selected influencer, we collect the initial sponsored video for each brand, along with the preceding non-sponsored video. We employ a quasi-experimental design that compares the video immediately above the sponsorship threshold to the one immediately below (Lee and Card 2007). This method is appropriate since the two popularity metrics under consideration are continuous. We only consider the first sponsored video within a brand-influencer pairing, as ongoing sponsorships may depend on an influencer's performance in previous sponsored videos, such as their filming style, in addition to the popularity metrics (Barker 2018). This approach enables the identification of influencers that meet the two popularity metrics but are not sponsored. However, it also runs the risk of generating a weak instrument problem that violates the RD assumption (Lee and Lemieux 2010).

In addition to gathering the first sponsored video and the non-sponsored video immediately preceding the sponsored video for each influencer, we also collect the influencer's follower count and average views over the three weeks before each relevant sponsored video upload. This timeframe is consistent with the typical gap between brand outreach and branded video uploads, as reported by a Neoreach sales manager, Aaron Layden.[9] The average views calculation is based on the method used by Mediakix (2019), which involves selecting the latest

[9] In interviews, we learned that brands care about three weeks of average views because it is a good predictor of the influencer's ability to attract views. Otherwise, brands risk recruiting an influencer with a lot of followers that rarely watch the influencer's content.

12 videos from the influencer, excluding the video with the highest and lowest views, and averaging the remaining ten video views. This approach enables us to capture the influencer's recent popularity metrics, which may better reflect the influencer's current reach and engagement levels, and thus their potential effectiveness as a sponsored influencer.

Certain influencers may meet the minimum follower count and average view thresholds set by a brand but may not publish any sponsored videos. The absence of sponsored content may be due to the influencer's personal reasons, such as a break from producing sponsored videos, or the brand's marketing budget constraints. The presence of influencers who surpass the thresholds but do not upload sponsored content causes imprecision in the discontinuities in our data. As such, we apply the Fuzzy RD approach to address this fuzziness.

OA 4.C. REPRESENTATIVENESS OF THE 10 BRANDS IN THE FRD MODEL

In the period spanning 2016 through 2018, we observed 21,033 sponsored videos from 3,296 unique brands. While the number of brands is relatively substantial, the distribution of sponsored videos is highly skewed to the right. This observation is documented in Table S26.

Table S26. Distribution of Brands by the Number of Sponsored Videos

Number of Sponsored Videos	1 video	2 videos	3 videos	4 videos	5 videos
Brand	1804 (54.73%)	495 (15.02%)	201 (6.10%)	152 (4.61%)	92 (2.79%)

Number of Sponsored Videos	6~10 videos	11~50 videos	51~100 videos	More than 100 videos	Total
Brand	221 (6.71%)	268 (8.13%)	30 (0.91%)	33 (1%)	21,033

A comprehensive understanding of market shares is crucial in determining the competitive landscape of an industry. However, due to the diverse range of brands involved in our study, we were unable to obtain a reliable data source for market shares. Instead, we narrow our focus to the top 10 brands, all of which sponsored more than 100 videos during the sample period and account for 12.26% of the total sponsored videos. Table S27 presents a detailed breakdown of the number of sponsorships from each of the top 10 brands. This approach allows us to provide a detailed analysis of the most prominent players in our sample and their respective sponsorship activities.

Table S27. Distribution of Sponsored Videos Among the RD Brands

Brand	Number of Sponsored Videos

Fabfitfun	397
Garnier	106
L'Oréal Paris	284
Maybelline NY	208
Neutrogena	168
Olay	155
Scentbird	378
Sephora	303
Ulta	210
Vanity Planet	370
Total	2,579 (12.26% of 21,033 sponsored videos created by 1,079 influencers)

Table S28 presents the count of distinct influencers sponsored by the top 10 brands during the sample period of 2016-2018. The 10 brands collectively sponsored over 50% of the influencers in our sample, indicating their significant influence in the market.

Table S28. Distribution of Sponsored Influencers Among the RD brands

Brand	Number of Influencers
Fabfitfun	184
Garnier	79
L'Oréal Paris	118
Maybelline New York	119
Neutrogena	108
Olay	93
Scentbird	235
Sephora	153
Ulta	112
Vanity Planet	218
Total	596 (55.24% of the 1,079 beauty influencers)

OA 4.D. FRD OPTIMAL FUNCTIONAL FORM SELECTION

In line with Lee and Lemieux (2010) and Jacob and Zhu (2012), we adopt F-statistics to conduct a comparison between models that include and exclude bin dummy covariates. To select the optimal RD model specification, we utilize the classic AIC. The outcomes of these analyses are presented in Table S29.

Table S29. RD Model Selection using Three Specifications

(Without Covariate)	Treatment Estimate	Standard Error	AIC	F-stat
Model 1) Linear	-0.012**	(0.005)	-3400.44	
Model 2) Quadratic	-0.014**	(0.005)	-3423.44	
Model 3) Cubic	-0.014**	(0.005)	**-3422.83**	
(With Bin Dummies)				
Model 1) Linear	-0.011*	(0.006)	-3478.28	3.016
Model 3) Quadratic	-0.012**	(0.006)	-3493.15	2.870
Model 5) Cubic	-0.027***	(0.006)	**-3604.97**	**4.344**
Notes. Bin dummies include the year and month dummies, featured brands, image feature labels, and six video format dummies.				

Among the three possible functional forms, we adopt the cubic model after examining its performance using the Akaike Information Criterion (AIC) and the F-statistic test. Specifically, we select the cubic model with bin dummy covariates, as it yields the lowest AIC value and a p-value of less than 0.0001 for the F-statistic, indicating the significance of the cubic interaction terms. The cubic functional form is expressed as:

$$f(X_{it} - c_m) = \phi_0 + \phi_1(X_{it} - c_m) + \phi_2(X_{it} - c_m)^2 + \phi_3(X_{it} - c_m)^3$$

where X_{it} is influencer *i*'s assignment variable at time *t,* and c_m is the cutoff (threshold) of firm m. Since we have two assignment variables (follower count and average views), our functional form is defined as:

$$\begin{aligned} f(X_{it} - c_m) = \phi_0 &+ \phi_1\left(follower\ count_{it} - c_m^f\right) + \phi_2\left(follower\ count_{it} - c_m^f\right)^2 \\ &+ \phi_3\left(follower\ count_{it} - c_m^f\right)^3 + \mu_1(average\ views_{it} - c_m^v) \\ &+ \mu_2(average\ views_{it} - c_m^v)^2 + \mu_3(average\ views_{it} - c_m^v)^3 \end{aligned}$$

In a robustness check, we repeat the estimations with the three functional forms (see results in OA 4.E).

OA 4.E. FRD MODEL COMPARISON USING THREE FUNCTIONAL FORMS

We conduct a robustness check of our proposed fuzzy regression discontinuity (FRD) model by testing three alternative functional forms: linear, quadratic, and cubic. The results are presented in Table S30 and are found to be qualitatively consistent with the main model. This indicates that our proposed FRD model is robust to different functional forms and reinforces the validity of our findings.

Table S30. FRD Model: Three Alternative Functional Forms

	Linear		Quadratic		Cubic	
Vocal characteristics	(1)		(2)		(3)	
Average loudness	0.0115**	(0.0050)	0.0117**	(0.0050)	0.0118**	(0.0050)
Average pitch	0.0156	(0.0157)	0.0157	(0.0157)	0.0155	(0.0158)
Loud variability	-0.0108**	(0.0055)	-0.0107**	(0.0055)	-0.0107*	(0.0055)
Pitch variability	-0.0142	(0.0108)	-0.0142	(0.0109)	-0.0142	(0.0109)
Talking duration	-0.0328**	(0.0139)	-0.0328**	(0.0139)	-0.0328**	(0.0139)
Sponsorship & Sponsored Vocal Characteristics						
Sponsorship	0.0060	(0.0146)	0.0058	(0.0147)	0.0051	(0.0147)
* *Average loudness*	-0.0176**	(0.0086)	-0.0179**	(0.0086)	-0.0188**	(0.0086)
* *Average pitch*	-0.0224	(0.0180)	-0.0225	(0.0181)	-0.0230	(0.0181)
* *Loud variability*	0.0141	(0.0090)	0.0141	(0.0090)	0.0147	(0.0090)
* *Pitch variability*	0.0268*	(0.0151)	0.0270*	(0.0151)	0.0272*	(0.0152)
* *Talking duration*	0.0343*	(0.0178)	0.0345*	(0.0179)	0.0356**	(0.0180)
Functional Forms of Assignment Variables						
(f_it-c^f_m)/10^6	0.0175	(0.0131)	0.0134	(0.0223)	-0.0491	(0.0495)
(v_it-c^v_m)/10^6	0.0004	(0.0061)	0.0009	(0.0030)	0.0301	(0.0197)
((f_it-c^f_m)/10^6)^2			0.0065	(0.0141)	-0.0026	(0.0016)
((v_it-c^v_m)/10^6)^2			-0.0007	(0.0012)	0.0208	(0.0249)
((f_it-c^f_m)/10^6)^3					-0.0057	(0.0066)
((v_it-c^v_m)/10^6)^3					0.0003	(0.0004)
Constant	1.2001***	(0.299)	1.200***	(0.301)	1.1847***	(0.301)
Observations (Individual)	2,514 (596)		2,514 (596)		2,514 (596)	
Fixed Effects	Individual, time, 30 brands, 30 product types, content					
R^2	0.5629		0.563		0.5637	

OA 4.F. FRD SENSITIVITY CHECK

In accordance with Jacob and Zhu (2012), we evaluate the sensitivity of the FRD model by gradually removing the outermost 1%, 5%, and 10% of the data. For our analysis, this equates to removing 25, 125, and 251 videos from our sample of 2,514 observations. We identify videos for removal by selecting those with the largest absolute value of the sum of the differences between assignment variables and cutoff values ($follower\ count_{it} - c_m^f$ and $average\ views_{it} - c_m^v$). The results of this sensitivity analysis are presented in Table S31. The findings reveal that the average loudness in sponsored videos remains statistically significant even after dropping out 1%, 5%, and 10% of the data. Moreover, the coefficients of interest exhibit qualitative consistency with the primary model across all three dropout specifications.

Table S31. FRD Sensitivity Check Results

	FRD 99% (1)	FRD 95% (2)	FRD 90% (3)

Vocal characteristics	Coef	SE	Coef	SE	Coef	SE
Average loudness	0.0126**	(0.0051)	0.0132**	(0.0053)	0.0126**	(0.0055)
Average pitch	0.0156	(0.0159)	0.0112	(0.0174)	0.0133	(0.0177)
Loud variability	-0.0116**	(0.0055)	-0.0103*	(0.0060)	-0.0109*	(0.0060)
Pitch variability	-0.0144	(0.0110)	-0.0128	(0.0120)	-0.0155	(0.0120)
Talking duration	-0.0344**	(0.0141)	-0.0297*	(0.0157)	-0.0323**	(0.0160)
Sponsorship & Sponsored Vocal Characteristics						
Sponsorship	0.0046	(0.0149)	0.0057	(0.0162)	-0.0017	(0.0175)
* *Average loudness*	-0.0210**	(0.0086)	-0.0202**	(0.0089)	-0.0182**	(0.0092)
* *Average pitch*	-0.0248	(0.0183)	-0.0191	(0.0208)	-0.0205	(0.0215)
* *Loud variability*	0.0153*	(0.0091)	0.0154	(0.0118)	0.0132	(0.0122)
* *Pitch variability*	0.0281*	(0.0152)	0.0268*	(0.0163)	0.0307*	(0.0168)
* *Talking duration*	0.0381**	(0.0182)	0.0335*	(0.0202)	0.0366*	(0.0210)
Constant	1.1295***	(0.3056)	1.2378***	(0.3262)	1.5434***	(0.3641)
Observations (Individual)	2,489 (589)		2,389 (587)		2,263 (577)	
Fixed Effects	Influencer, year-month, and content control variables (visual, verbal, textual, popularity and ad spending data)					
R^2	0.5643		0.5658		0.5803	
Notes. Table entries are coefficients, and the robust standard errors are in parentheses. *p < 0.10; **p < 0.05; ***p < 0.01.						

OA SECTION 5. FULL MODEL RESULTS

OA 5.A. FULL RESULTS OF TABLE 8

The complete set of coefficients for the estimation of Equation (2) can be found in Table S32. Please note that the abbreviated results are also presented in Table 8 in the main paper for ease of reference.

Table S32. Change in Vocal Characteristics in Sponsored Videos (Relative to Non-Sponsored Videos)

	(1)	(2)	(3)	(4)	(5)
	Average Loudness	**Average Pitch**	**Loudness Variability**	**Pitch Variability**	**Talking Duration**
Sponsorship	**-0.069*****	**0.063*****	**0.026****	**-0.040*****	**0.097*****
	(0.010)	**(0.010)**	**(0.012)**	**(0.010)**	**(0.009)**
Other Content Control Variables					
Duration	-0.274***	-0.081***	0.001	-0.048***	-0.008
	(0.016)	(0.013)	(0.019)	(0.013)	(0.013)
View count	-0.014	-0.002	0.033*	-0.020	0.003
	(0.013)	(0.013)	(0.018)	(0.012)	(0.013)
Like count	-0.020	0.057***	-0.059***	0.036**	0.047***
	(0.018)	(0.017)	(0.022)	(0.016)	(0.017)
Dislike count	0.016*	-0.039***	0.030***	-0.007	-0.039***
	(0.010)	(0.009)	(0.012)	(0.009)	(0.009)
Comment count	0.050***	-0.010	0.010	-0.012	-0.003
	(0.010)	(0.010)	(0.011)	(0.010)	(0.009)
Title length	-0.015	0.031*	0.002	0.016	0.025
	(0.016)	(0.016)	(0.017)	(0.017)	(0.015)
Description length	0.067***	0.070***	-0.015	-0.017	0.085***
	(0.016)	(0.013)	(0.017)	(0.013)	(0.012)
Tenure	-0.059	0.005	-0.019	-0.066***	0.046*
	(0.039)	(0.031)	(0.029)	(0.024)	(0.025)
Followers	-0.049**	-0.003	-0.034	-0.012	0.054**
	(0.025)	(0.021)	(0.041)	(0.019)	(0.022)
Account total views	-0.011	0.004	-0.015	-0.018	0.031
	(0.023)	(0.022)	(0.017)	(0.021)	(0.019)
Instagram followers	0.024	0.003	-0.046*	0.005	0.014
	(0.018)	(0.018)	(0.025)	(0.016)	(0.016)
Avg. diff. in followers	-0.004	0.011	0.006	-0.006	0.017***

	(0.006)	(0.007)	(0.005)	(0.007)	(0.006)
Avg. diff. in total views	-0.001	0.001	-0.005	-0.002	0.002
	(0.008)	(0.008)	(0.013)	(0.007)	(0.008)
Avg. diff. in Instagram followers	-0.0004	0.0003	0.003	-0.002	-0.001
	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)
Format - GRWM	0.044***	-0.013	-0.064***	0.106***	-0.081***
	(0.015)	(0.015)	(0.017)	(0.016)	(0.014)
Format - haul	0.073***	0.031**	0.048***	-0.029**	0.049***
	(0.014)	(0.013)	(0.014)	(0.013)	(0.012)
Format - review	-0.048***	0.029**	0.027*	-0.028**	0.046***
	(0.012)	(0.013)	(0.015)	(0.014)	(0.012)
Format - routine	-0.069***	-0.012	0.027*	-0.038***	0.016
	(0.012)	(0.013)	(0.016)	(0.014)	(0.012)
Format - tutorial	-0.037***	0.039***	0.033**	-0.042***	0.070***
	(0.014)	(0.013)	(0.016)	(0.013)	(0.013)
Format - vlog	0.098***	-0.096***	-0.092***	0.205***	-0.220***
	(0.015)	(0.015)	(0.014)	(0.015)	(0.014)
Verbal sophistication 1gram	0.160***	-0.028	0.059	-0.076***	-0.033
	(0.040)	(0.026)	(0.066)	(0.024)	(0.032)
Verbal sophistication 2gram	-0.219***	-0.029	0.099*	-0.065*	0.048
	(0.056)	(0.042)	(0.052)	(0.035)	(0.038)
Verbal sophistication 3gram	-0.012	0.090***	-0.092**	0.016	0.086***
	(0.029)	(0.027)	(0.039)	(0.024)	(0.027)
Verbal concreteness	-0.594***	0.174	-0.449*	0.506***	0.084
	(0.196)	(0.160)	(0.236)	(0.138)	(0.154)
Verbal topic 0	-0.085***	0.048**	-0.034	0.024	0.042**
	(0.020)	(0.020)	(0.022)	(0.020)	(0.018)
Verbal topic 1	0.061	0.020	0.007	0.120***	-0.076**
	(0.039)	(0.041)	(0.036)	(0.041)	(0.035)
Verbal topic 2	-0.102***	-0.032	-0.024	-0.038	0.055
	(0.037)	(0.048)	(0.045)	(0.049)	(0.046)
Verbal topic 3	-0.051	0.016	0.063**	-0.095***	0.085***
	(0.031)	(0.033)	(0.027)	(0.034)	(0.028)
Verbal topic 4	-0.100***	0.033	0.022	-0.045	0.110***
	(0.026)	(0.037)	(0.037)	(0.039)	(0.032)
Verbal topic 5	-0.111***	0.056	-0.018	0.107**	-0.024
	(0.029)	(0.051)	(0.063)	(0.044)	(0.047)
Verbal topic 6	0.002	0.085	-0.085*	-0.067	0.108**
	(0.030)	(0.062)	(0.046)	(0.062)	(0.044)

Verbal topic 7	-0.199***	-0.042	-0.031	-0.088	0.101**
	(0.039)	(0.050)	(0.050)	(0.054)	(0.045)
Verbal topic 8	0.016	-0.044	0.137	0.016	-0.103
	(0.035)	(0.059)	(0.113)	(0.062)	(0.066)
Verbal topic 9	-0.100**	0.087	0.043	-0.078	0.128**
	(0.051)	(0.066)	(0.051)	(0.064)	(0.055)
Age	-0.124**	0.092*	0.008	-0.093*	0.171***
	(0.051)	(0.053)	(0.052)	(0.054)	(0.047)
Emotion - angry	-0.033	0.057*	-0.075***	-0.034	0.093***
	(0.031)	(0.032)	(0.024)	(0.032)	(0.029)
Emotion - disgust	0.010	0.078	-0.021	-0.060	0.118**
	(0.048)	(0.057)	(0.042)	(0.062)	(0.048)
Emotion - fear	0.035	-0.062**	0.039	-0.109***	0.025
	(0.026)	(0.028)	(0.026)	(0.028)	(0.025)
Emotion - happy	0.003	0.006	0.026	-0.020	0.016
	(0.016)	(0.016)	(0.017)	(0.016)	(0.015)
Emotion - sad	0.033	-0.007	0.025	-0.018	-0.010
	(0.022)	(0.023)	(0.026)	(0.023)	(0.021)
Emotion - surprise	-0.021	0.037	0.015	-0.020	0.053**
	(0.026)	(0.029)	(0.024)	(0.031)	(0.026)
Race - asian	0.025	-0.036	-0.025	-0.122***	0.105***
	(0.038)	(0.040)	(0.038)	(0.039)	(0.036)
Race - black	0.001	-0.040	0.062	-0.165***	0.113***
	(0.054)	(0.045)	(0.055)	(0.045)	(0.044)
Race - indian	0.040	-0.023	-0.022	-0.076	0.027
	(0.070)	(0.077)	(0.072)	(0.078)	(0.070)
Race - latino_hispanic	-0.056	0.087	-0.025	-0.089	0.213***
	(0.061)	(0.061)	(0.059)	(0.063)	(0.056)
Race - middle_eastern	0.076	-0.076	0.069	-0.085	-0.019
	(0.050)	(0.050)	(0.048)	(0.053)	(0.044)
Facial attractiveness	0.086	0.023	-0.181**	0.284***	-0.154**
	(0.060)	(0.063)	(0.079)	(0.063)	(0.060)
Makeup heaviness	0.010	-0.551***	0.022	0.815***	-1.260***
	(0.231)	(0.197)	(0.262)	(0.195)	(0.193)
Object – top/t-shirt/sweatshirt	0.076	0.036	0.014	-0.013	0.056*
	(0.055)	(0.036)	(0.033)	(0.038)	(0.032)
Object – pants	0.046	0.035	-0.036	0.091	-0.013
	(0.100)	(0.080)	(0.064)	(0.079)	(0.070)
Object – skirt	-0.094	-0.565**	0.154	0.133	-0.569***

	(0.180)	(0.228)	(0.144)	(0.146)	(0.196)
Object – dress	0.029**	-0.001	-0.034**	0.022*	-0.013
	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)
Object - glasses	-0.075	-0.002	0.017	-0.168**	0.139***
	(0.053)	(0.054)	(0.043)	(0.065)	(0.050)
Object – hat	-0.369	0.038	0.012	0.112	0.004
	(0.268)	(0.221)	(0.097)	(0.140)	(0.258)
Object – watch	-0.146	0.391**	-0.431	-0.261**	0.749***
	(0.148)	(0.161)	(0.312)	(0.128)	(0.182)
Object – belt	-0.167	1.441**	-0.913	0.415*	1.029*
	(0.187)	(0.577)	(0.707)	(0.248)	(0.539)
Object – tights/stockings	0.419	0.130	0.564	0.108	-0.118
	(0.357)	(0.159)	(0.411)	(0.144)	(0.215)
Object – shoe	0.051***	0.002	-0.009	0.008	-0.024
	(0.019)	(0.017)	(0.018)	(0.017)	(0.016)
Object – bag/wallet	-0.038	-0.185	-0.062	-0.270	0.009
	(0.138)	(0.256)	(0.105)	(0.241)	(0.164)
Object – collar	0.897*	-0.190	-0.123	0.337**	-0.375*
	(0.512)	(0.206)	(0.117)	(0.136)	(0.194)
Object – lapel	-0.293**	-0.168	0.308	0.107	-0.326*
	(0.140)	(0.175)	(0.350)	(0.241)	(0.183)
Object – sleeve	0.025**	-0.001	-0.016	0.028**	-0.034***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)
Object – pocket	-0.196	-0.139	-0.021	0.212*	-0.295
	(0.240)	(0.216)	(0.153)	(0.115)	(0.217)
Object – neckline	-0.011	0.009	0.001	-0.012	0.009
	(0.011)	(0.012)	(0.011)	(0.012)	(0.010)
Foundation 1 (G) value	0.045	0.094	-0.011	-0.328***	0.337***
	(0.070)	(0.069)	(0.098)	(0.066)	(0.068)
Foundation 2 (G) value	-0.024	0.027	0.054	-0.107	0.191***
	(0.074)	(0.074)	(0.071)	(0.075)	(0.067)
Foundation 3 (G) value	-0.041	0.030	0.132*	-0.149**	0.131**
	(0.065)	(0.064)	(0.072)	(0.065)	(0.059)
Blush 1 (G) value	0.138	0.040	0.028	-0.114	0.130
	(0.120)	(0.105)	(0.098)	(0.105)	(0.096)
Blush 2 (G) value	0.238	-0.158	-0.043	-0.237	0.020
	(0.179)	(0.154)	(0.131)	(0.151)	(0.142)
Blush 3 (G) value	-0.247	-0.075	0.144	0.201	-0.204
	(0.215)	(0.219)	(0.167)	(0.209)	(0.190)
Blush 4 (G) value	0.180	-0.017	-0.009	0.114	-0.131

	(0.229)	(0.264)	(0.203)	(0.260)	(0.226)
Blush 5 (G) value	-0.493***	0.178	-0.124	0.106	0.038
	(0.165)	(0.177)	(0.162)	(0.180)	(0.161)
Lip (G) value	-0.003	-0.083	-0.120**	0.119**	-0.176***
	(0.050)	(0.052)	(0.048)	(0.052)	(0.046)
Lipliner (G) value	0.044	-0.032	0.105	-0.136*	0.041
	(0.080)	(0.080)	(0.084)	(0.080)	(0.072)
Eyeshadow1 (G) value	-0.055	-0.060	0.041	-0.138**	0.056
	(0.054)	(0.055)	(0.060)	(0.056)	(0.051)
Eyeshadow2 (G) value	-0.247**	0.391***	0.208*	-0.229*	0.525***
	(0.100)	(0.120)	(0.114)	(0.128)	(0.105)
Eyeshadow3 (G) value	0.330**	0.018	-0.196	0.410**	-0.301**
	(0.150)	(0.175)	(0.165)	(0.188)	(0.154)
Eyeshadow4 (G) value	-0.188	-0.441**	0.088	-0.308	-0.212
	(0.188)	(0.211)	(0.190)	(0.223)	(0.186)
Eyeshadow5 (G) value	-0.127	0.272*	0.109	0.038	0.238*
	(0.131)	(0.145)	(0.152)	(0.153)	(0.133)
Eyeliner (G) value	0.089	-0.119**	-0.181***	0.278***	-0.294***
	(0.056)	(0.056)	(0.058)	(0.057)	(0.052)
Eyebrow (G) value	0.141***	-0.083	-0.171***	0.072	-0.064
	(0.047)	(0.051)	(0.064)	(0.051)	(0.049)
NIMA score average	0.370***	-0.047**	-0.073***	0.218***	-0.252***
	(0.027)	(0.021)	(0.027)	(0.020)	(0.020)
Monthly ad. spending	0.001	0.002	-0.002	0.0001	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	2.714***	-0.933*	1.208	1.178*	-2.913***
	(0.548)	(0.557)	(0.862)	(0.704)	(0.552)
Observations (Individual)	32,920 (1,017)	32,920 (1,017)	32,920 (1,017)	32,920 (1,017)	32,920 (1,017)
Fixed Effects	Individual, year-month (time), 30 brands, 30 product types				
R^2	0.3461	0.3212	0.1947	0.3043	0.4254

OA 5.B. SPONSORSHIP COEFFICIENTS OF THE FOUR VOCAL FEATURES

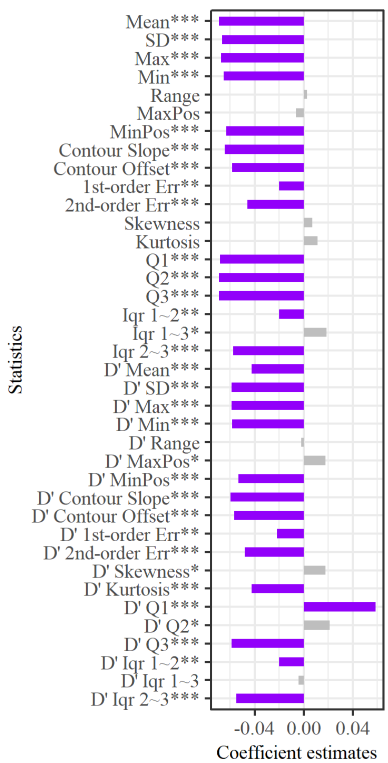
Thus far, our analyses have focused on the mean values of the four primary vocal characteristics, namely loudness, pitch, loudness variability, and taking duration, along with the standard deviation of pitch variability (as described in OA 2.B). However, openSMILE computes a total of 19 statistics and their first-order derivatives for each primary vocal feature, resulting in a comprehensive set of 152 vocal statistics: 4 features x (19 stats + 19 first-order derivatives).

These statistics are defined in Table S6, as outlined in OA 2.B, providing a comprehensive overview of the computational measures employed in our investigation.

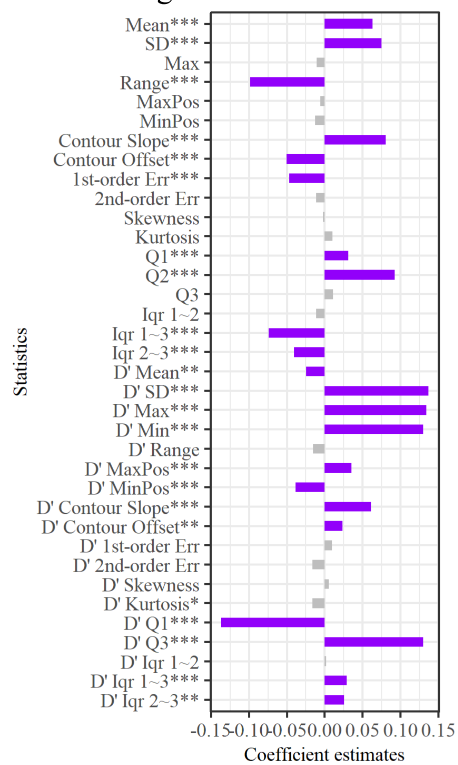
The extended plots of the sponsorship coefficients of the 150 vocal statistics (2 dropped out of 152 total statistics due to no changes), including the first-order derivatives denoted by variables starting with D', are presented in Figure S8. Notably, 69% of these coefficients (104 out of 150) exhibit statistical significance, lending support to our assertion that influencers purposefully modulate their vocal characteristics in sponsored videos compared to non-sponsored videos. Additionally, our findings reveal negative coefficients for various measures of average loudness, including maximum, range, standard deviation, and mean, aligning with our main result that influencers tend to decrease their average loudness in sponsored videos. These results provide further empirical evidence of the vocal manipulation strategies employed by influencers in the context of sponsored content.

Figure S8. Sponsorship Coefficients of 152 Four Vocal Statistics

Average Loudness Statistics

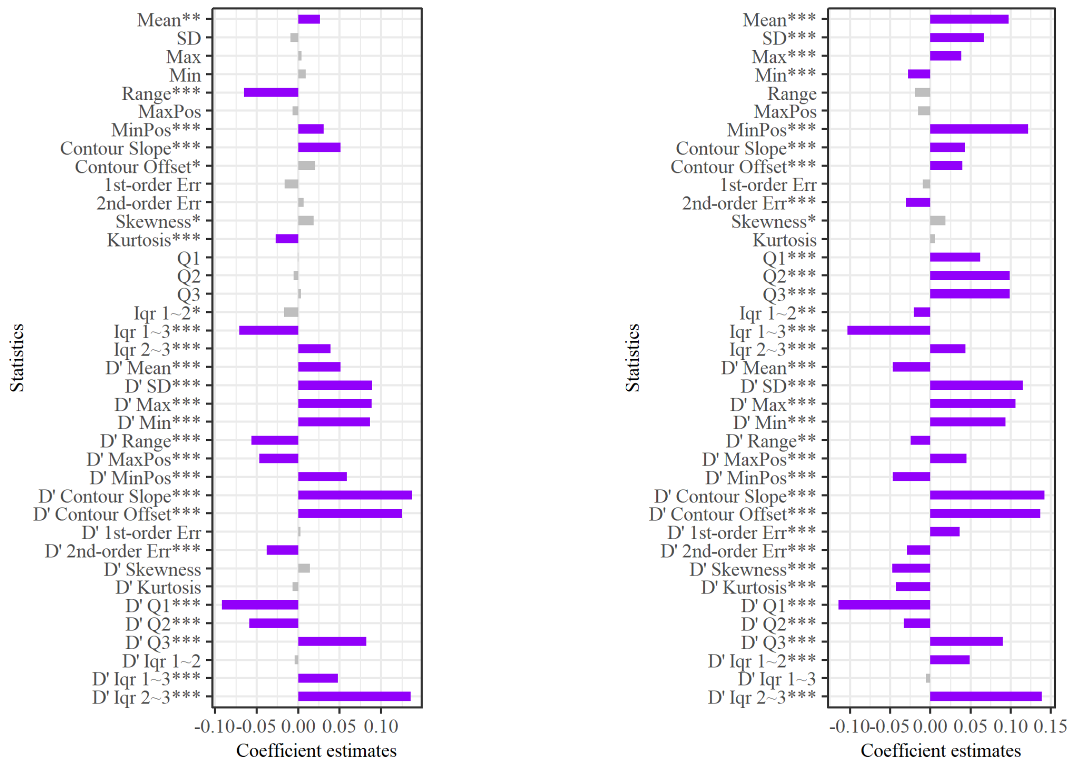


Average Pitch Statistics



Loudness Variability Statistics

Talking Duration Statistics



Notes. The purple bars indicate significant coefficients. MaxPos & MinPos are the absolute frame positions of the maximum and minimum values. Contour slope & offset are the slope and offset of a linear approximation of the vocal feature contour. 1st-order and 2nd-order errs are the linear and quadratic errors of the difference between the linear approximation of the vocal feature and its contour. Skewness and kurtosis are the 3rd- and 4th-order moments of the vocal feature. Q1, Q2, and Q3 are the 1^{st}, 2^{nd}, and 3^{rd} quantiles. Iqr 1~2, 2~3, and 1~3 are the interquartile ranges. D’ indicates the first derivative.

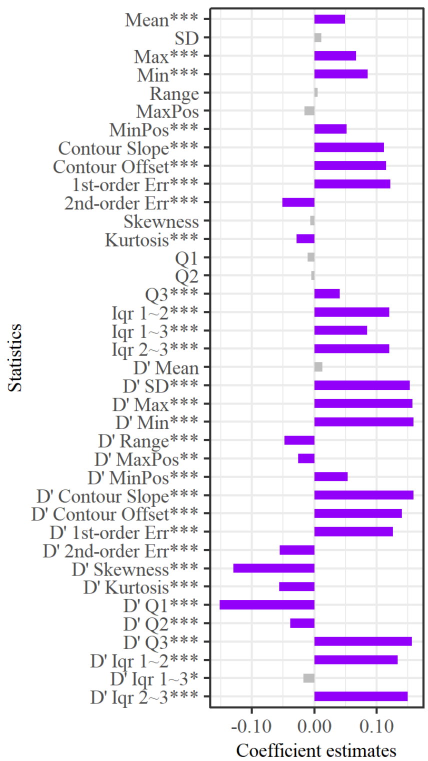
OA 5.C. SPONSORSHIP COEFFICIENTS OF SPECTRAL FEATURE STATISTICS

In addition to the four interpretable vocal characteristics, the literature on Affective Computing considers non-interpretable spectral features. The current study extends the analysis to two primary spectral vocal features, as illustrated in Figure S9. The first feature is line spectral pairs (LSP) frequency, which is a commonly used non-interpretable spectral feature of human speech. LSP frequency represents linear prediction coefficients for voice filter stability and vocal sound representational efficiency (McLoughlin 2008). The second feature is the Mel Frequency Cepstral Coefficient (MFCC), which is one of the most popular feature extraction techniques in Automatic Speech Recognition (ASR). The MFCC is based on the frequency domain using the Mel scale, which is based on the human ear scale (Dave 2013). The extended plots of the coefficients of these two features are shown in Figure S9.

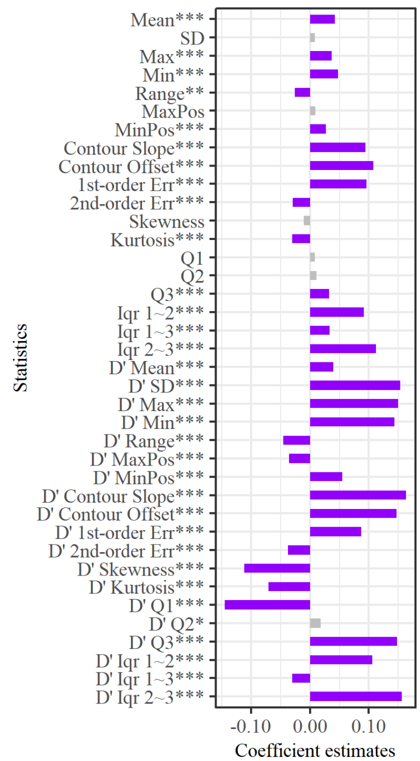
We find significant sponsorship coefficients for 570 of the 760 spectral vocal feature statistics (75%), supporting our conclusion that influencers strategically change their vocal characteristics in sponsored videos relative to non-sponsored videos.

Figure S9. Sponsorship Coefficients of 760 Spectral Vocal Feature Statistics

Line Spectral Pairs (LSP) Frequency (0)

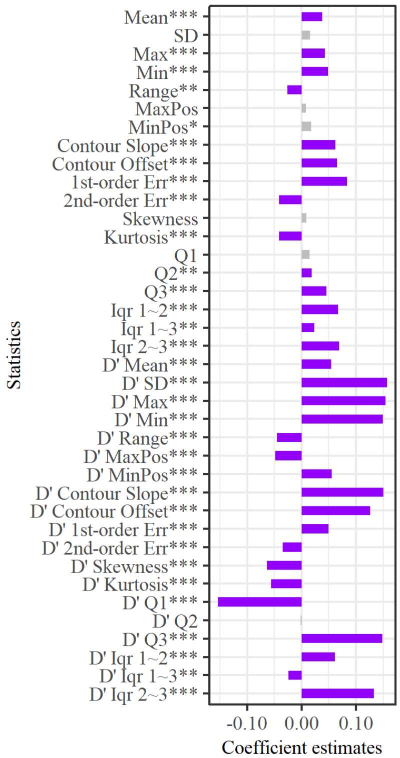


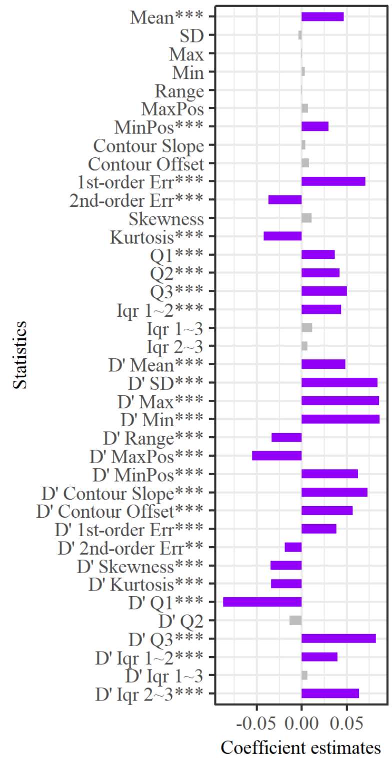
LSP Frequency (1)



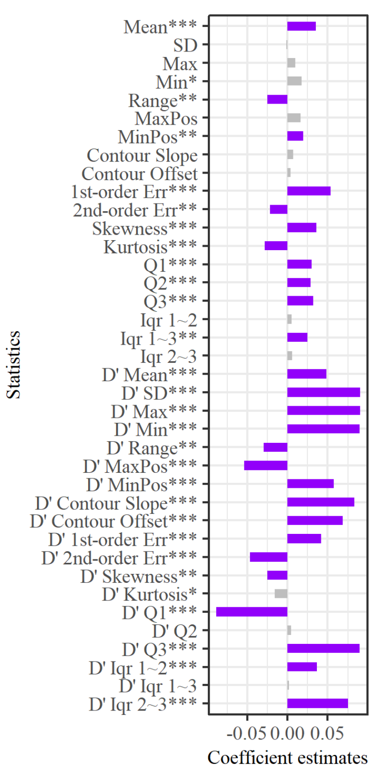
LSP Frequency (2)

LSP Frequency (3)

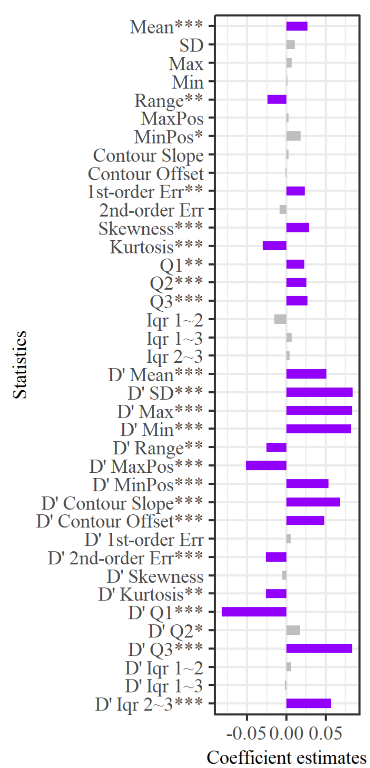




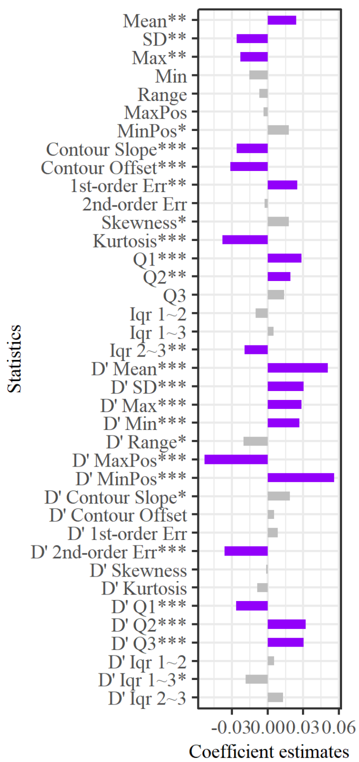
LSP Frequency (4)



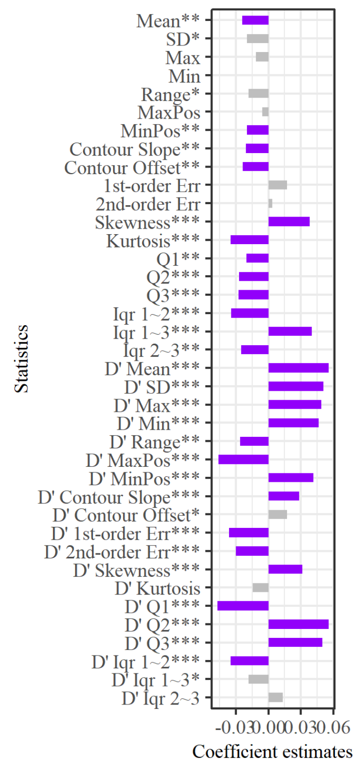
LSP Frequency (5)



LSP Frequency (6)

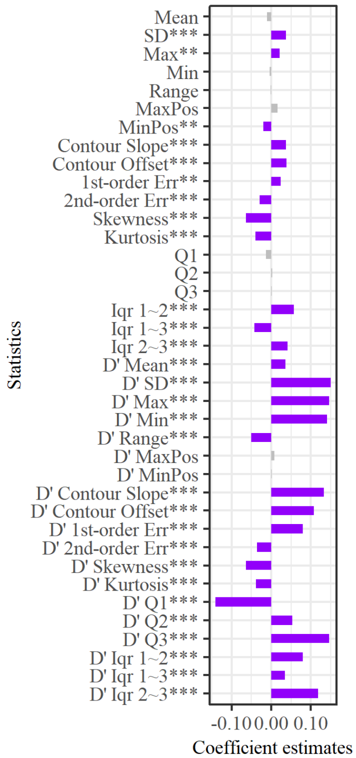


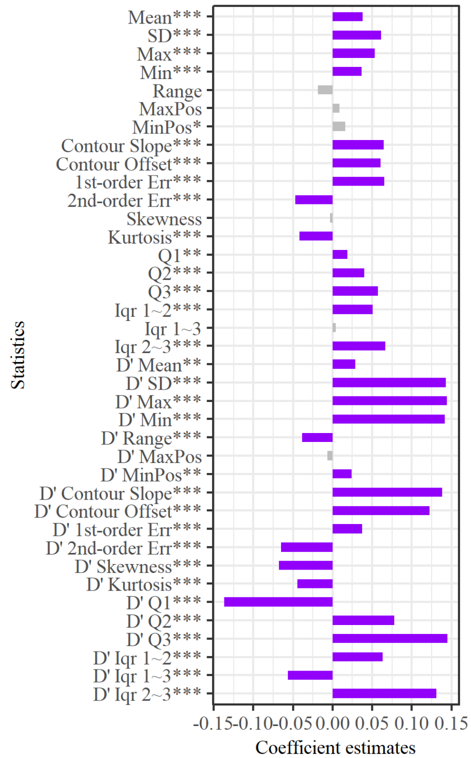
LSP Frequency (7)



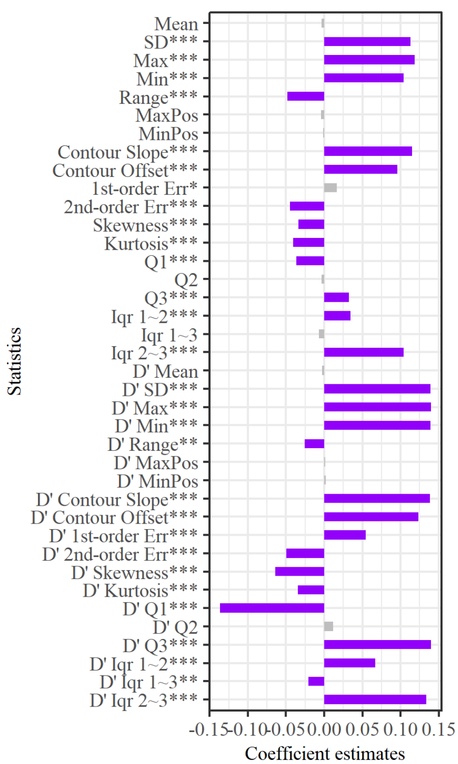
Mel Frequency Cepstral Coef. (MFCC) (1)

MFCC (2)

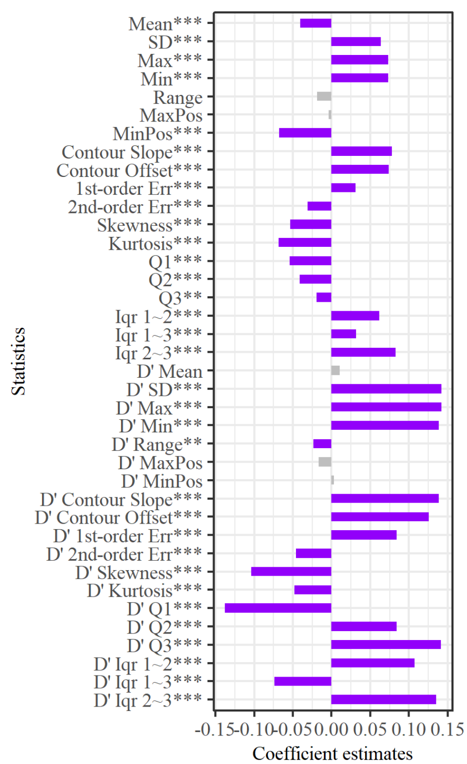

Mean
SD***
Max**
Min
Range
MaxPos
MinPos**
Contour Slope***
Contour Offset***
1st-order Err**
2nd-order Err***
Skewness***
Kurtosis***
Q1
Q2
Q3
Iqr 1~2***
Iqr 1~3***
Iqr 2~3***
D' Mean***
D' SD***
D' Max***
D' Min***
D' Range***
D' MaxPos
D' MinPos
D' Contour Slope***
D' Contour Offset***
D' 1st-order Err***
D' 2nd-order Err***
D' Skewness***
D' Kurtosis***
D' Q1***
D' Q2***
D' Q3***
D' Iqr 1~2***
D' Iqr 1~3***
D' Iqr 2~3***
Statistics
-0.10 0.00 0.10
Coefficient estimates


Mean***
SD***
Max***
Min***
Range
MaxPos
MinPos*
Contour Slope***
Contour Offset***
1st-order Err***
2nd-order Err***
Skewness
Kurtosis***
Q1**
Q2***
Q3***
Iqr 1~2***
Iqr 1~3
Iqr 2~3***
D' Mean**
D' SD***
D' Max***
D' Min***
D' Range***
D' MaxPos
D' MinPos**
D' Contour Slope***
D' Contour Offset***
D' 1st-order Err***
D' 2nd-order Err***
D' Skewness***
D' Kurtosis***
D' Q1***
D' Q2***
D' Q3***
D' Iqr 1~2***
D' Iqr 1~3***
D' Iqr 2~3***
Statistics
-0.15 -0.10 -0.05 0.00 0.05 0.10 0.15
Coefficient estimates

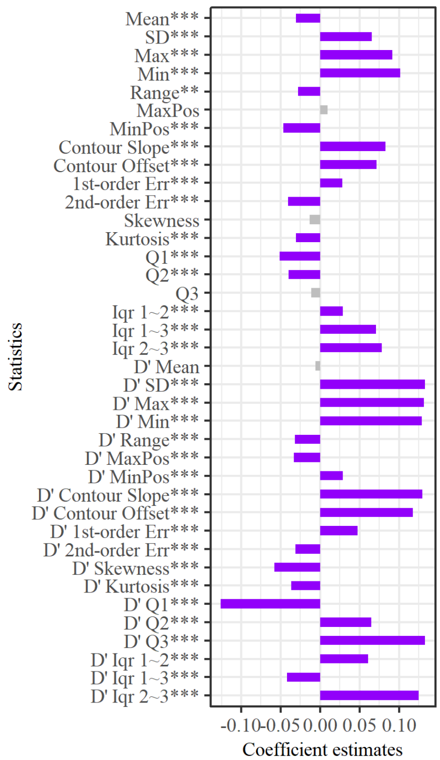
MFCC (3)


Mean
SD***
Max***
Min***
Range***
MaxPos
MinPos
Contour Slope***
Contour Offset***
1st-order Err*
2nd-order Err***
Skewness***
Kurtosis***
Q1***
Q2
Q3***
Iqr 1~2***
Iqr 1~3
Iqr 2~3***
D' Mean
D' SD***
D' Max***
D' Min***
D' Range**
D' MaxPos
D' MinPos
D' Contour Slope***
D' Contour Offset***
D' 1st-order Err***
D' 2nd-order Err***
D' Skewness***
D' Kurtosis***
D' Q1***
D' Q2
D' Q3***
D' Iqr 1~2***
D' Iqr 1~3**
D' Iqr 2~3***
Statistics
-0.15 -0.10 -0.05 0.00 0.05 0.10 0.15
Coefficient estimates

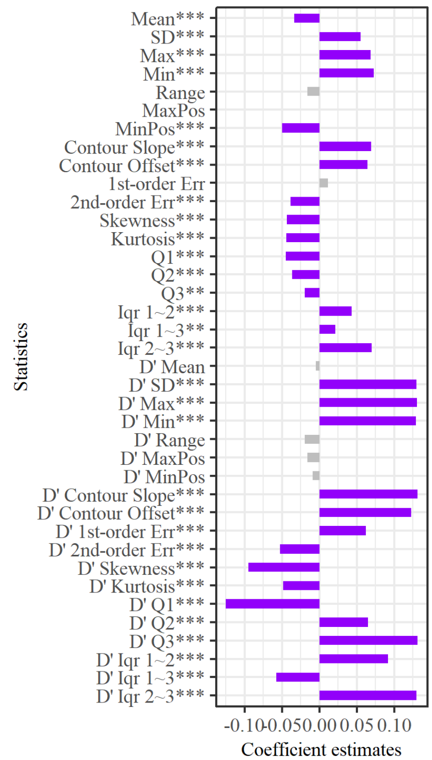
MFCC (4)


Mean***
SD***
Max***
Min***
Range
MaxPos
MinPos***
Contour Slope***
Contour Offset***
1st-order Err***
2nd-order Err***
Skewness***
Kurtosis***
Q1***
Q2***
Q3**
Iqr 1~2***
Iqr 1~3***
Iqr 2~3***
D' Mean
D' SD***
D' Max***
D' Min***
D' Range**
D' MaxPos
D' MinPos
D' Contour Slope***
D' Contour Offset***
D' 1st-order Err***
D' 2nd-order Err***
D' Skewness***
D' Kurtosis***
D' Q1***
D' Q2***
D' Q3***
D' Iqr 1~2***
D' Iqr 1~3***
D' Iqr 2~3***
Statistics
-0.15 -0.10 -0.05 0.00 0.05 0.10 0.15
Coefficient estimates

MFCC (5)

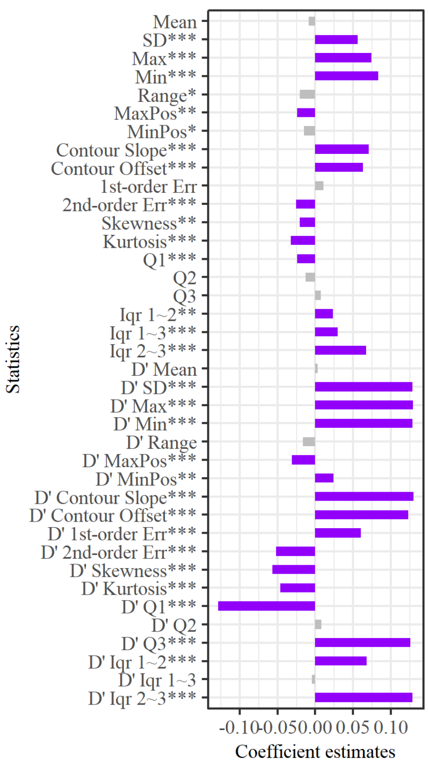


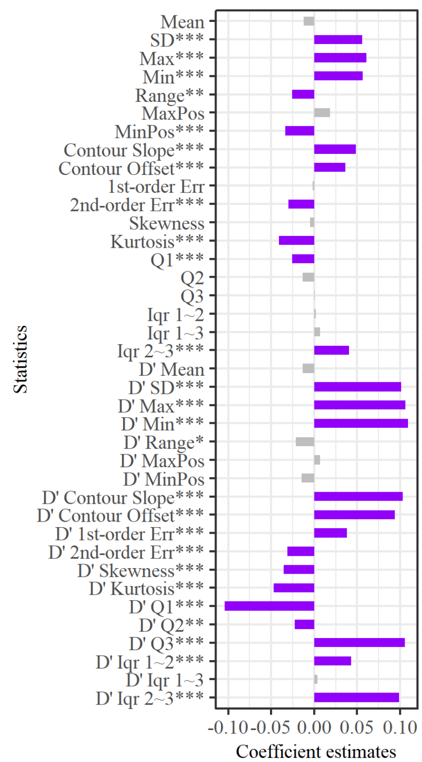
MFCC (6)



MFCC (7)

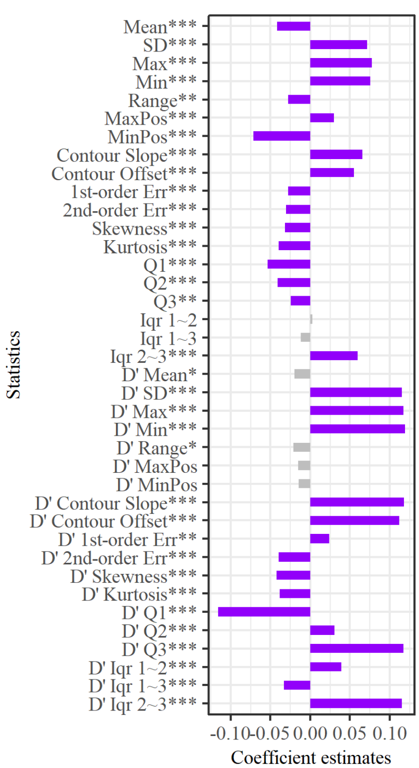
MFCC (8)

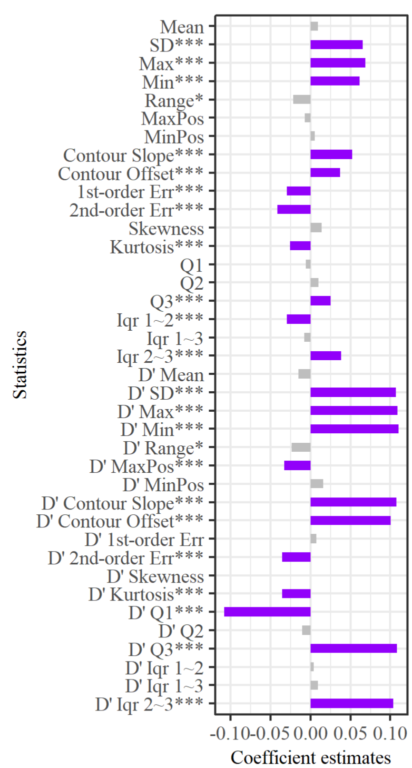


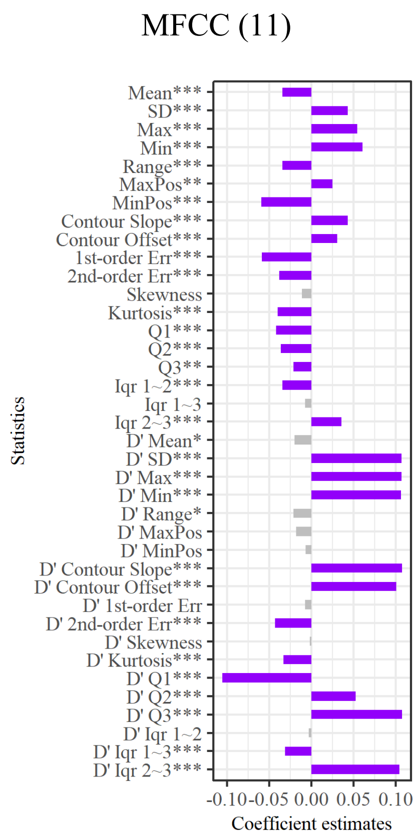


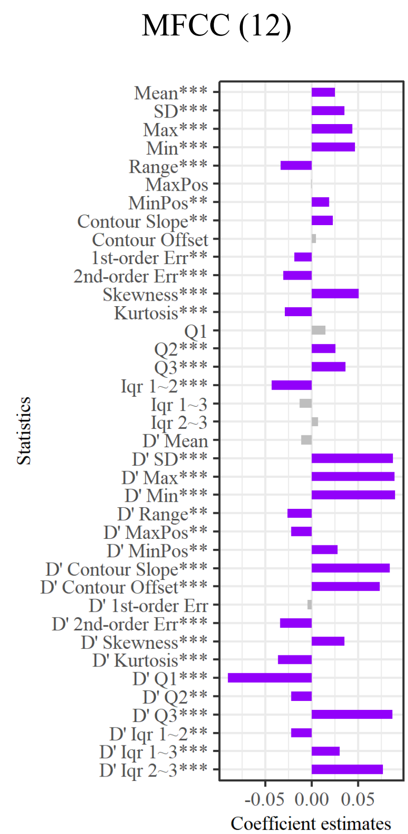
MFCC (9)

MFCC (10)









OA 5.D. FULL RESULTS OF TABLE 9

Table S33 provides all coefficients for the estimation of the IV-PSM model without interactions between sponsorship and the content control variables. (The abbreviated results appear in Table 9 in the main paper.)

Table S33. Effects of Sponsorship, Vocal Characteristics, and Interactions on Sentiment: IV Model

Variable	Coef.	SE
Average loudness	0.003**	(0.001)
Average pitch	0.006	(0.004)
Loudness variability	-0.002	(0.002)
Pitch variability	-0.003	(0.003)
Talking duration	-0.009**	(0.004)
Sponsorship & Sponsored Vocal Characteristics		
Sponsorship	-0.006**	(0.003)
Sponsorship * Average loudness	-0.005**	(0.002)
Sponsorship * Average pitch	-0.007	(0.007)

Sponsorship * Loudness variability	-0.003	(0.003)
Sponsorship * Pitch variability	0.001	(0.006)
Sponsorship * Talking duration	0.005	(0.007)
Other Content Control Variables		
Duration	0.003	(0.002)
View count	-0.025***	(0.002)
Like count	0.033***	(0.003)
Dislike count	-0.017***	(0.002)
Comment count	-0.008***	(0.002)
Title length	-0.004*	(0.002)
Description length	0.010***	(0.002)
Tenure	0.003	(0.004)
Followers	-0.004	(0.003)
Account total views	-0.001	(0.003)
Instagram followers	-0.002	(0.002)
Avg. diff. in followers	-0.001	(0.001)
Avg. diff. in total views	-0.0002	(0.001)
Avg. diff in Instagram followers	0.0003	(0.001)
Format - GRWM	0.007***	(0.002)
Format - haul	0.006***	(0.002)
Format - review	-0.002	(0.002)
Format - routine	-0.005***	(0.002)
Format - tutorial	-0.002	(0.002)
Format - vlog	0.001	(0.002)
Verbal sophistication 1gram	-0.008***	(0.003)
Verbal sophistication 2gram	0.007	(0.005)
Verbal sophistication 3gram	0.004	(0.004)
Verbal concreteness	0.009	(0.018)
Verbal topic 0	-0.0002	(0.003)
Verbal topic 1	0.003	(0.006)
Verbal topic 2	0.010	(0.008)
Verbal topic 3	0.017***	(0.005)
Verbal topic 4	-0.018***	(0.006)
Verbal topic 5	-0.014**	(0.006)
Verbal topic 6	0.003	(0.007)
Verbal topic 7	-0.019**	(0.009)
Verbal topic 8	0.019**	(0.008)
Verbal topic 9	-0.019*	(0.010)
Age	0.003	(0.007)
Emotion - angry	-0.006	(0.005)

Emotion - disgust	0.012	(0.009)
Emotion - fear	0.001	(0.004)
Emotion - happy	0.007***	(0.002)
Emotion - sad	0.003	(0.003)
Emotion - surprise	-0.002	(0.004)
Race - asian	0.013**	(0.006)
Race - black	0.009	(0.007)
Race - indian	0.018*	(0.010)
Race - latino_hispanic	0.011	(0.009)
Race - middle_eastern	0.003	(0.007)
Facial attractiveness	-0.002	(0.009)
Makeup heaviness	-0.014	(0.029)
Object – top/t-shirt/sweatshirt	0.006	(0.005)
Object – pants	0.008	(0.010)
Object – skirt	-0.005	(0.054)
Object – dress	0.001	(0.002)
Object - glasses	-0.017	(0.011)
Object – hat	0.052	(0.040)
Object – watch	-0.035*	(0.018)
Object – belt	0.031	(0.044)
Object – tights/stockings	-0.022	(0.021)
Object – shoe	0.005**	(0.002)
Object – bag/wallet	-0.039	(0.038)
Object – collar	0.013	(0.021)
Object – lapel	0.026	(0.028)
Object – sleeve	-0.002	(0.002)
Object – pocket	0.039	(0.026)
Object – neckline	-0.00001	(0.002)
Foundation1_G	-0.015	(0.009)
Foundation2_G	0.006	(0.011)
Foundation3_G	0.013	(0.009)
Blush1_G	0.012	(0.015)
Blush2_G	0.003	(0.022)
Blush3_G	-0.037	(0.031)
Blush4_G	0.056	(0.038)
Blush5_G	-0.027	(0.026)
Lip_G	-0.004	(0.007)
Lipliner_G	0.001	(0.011)
Eyeshadow1_G	-0.006	(0.008)
Eyeshadow2_G	0.022	(0.017)

Eyeshadow3_G	-0.003	(0.025)
Eyeshadow4_G	-0.050*	(0.030)
Eyeshadow5_G	0.033	(0.022)
Eyeliner_G	-0.013	(0.009)
Eyebrow_G	0.003	(0.007)
Average NIMA score	-0.002	(0.003)
Monthly ad. spending	0.0001	(0.0003)
Constant	0.761***	(0.085)
Observations (individual)	32,920 (1,017)	
Fixed Effects	Individual, year-month (time), 30 brands, 30 product types	
R^2	0.3355	
Note:*p<0.1;**p<0.05;***p<0.01		

OA 5.E. FULL RESULTS OF TABLE 10

Table S34 provides all coefficients for the estimations of the FRD and TWFE models. (The abbreviated results appear in Table 10 in the main paper.)

Table S34. Effects of Sponsorship, Vocal Characteristics, and Interactions on Sentiment: FRD & TWFE Models

	DV: Consumer Sentiment			
	FRD (1)		TWFE (2)	
Variable	Coef.	SE	Coef.	SE
Average loudness	**0.0118****	**(0.0050)**	**0.002****	**(0.001)**
Average pitch	**0.0155**	**(0.0158)**	**0.002**	**(0.002)**
Loudness variability	**-0.0107***	**(0.0055)**	**-0.002****	**(0.001)**
Pitch variability	**-0.0142**	**(0.0109)**	**0.001**	**(0.001)**
Talking duration	**-0.0328****	**(0.0139)**	**-0.005*****	**(0.002)**
Sponsorship & Sponsored Vocal Characteristics				
Sponsorship	**0.0051**	**(0.0147)**	**-0.014*****	**(0.002)**
Sponsorship * Average loudness	**-0.0188****	**(0.0086)**	**-0.004****	**(0.002)**
Sponsorship * Average pitch	**-0.0230**	**(0.0181)**	**-0.004**	**(0.004)**
Sponsorship * Loudness variability	**0.0147**	**(0.0090)**	**-0.003**	**(0.002)**
Sponsorship * Pitch variability	**0.0272***	**(0.0152)**	**-0.0001**	**(0.003)**
Sponsorship * Talking duration	**0.0356****	**(0.0180)**	**0.003**	**(0.004)**
Other Content Control Variables				
Duration	0.0033	(0.0096)	0.007***	(0.002)

View count	-0.0145*	(0.0084)	-0.032***	(0.002)
Like count	0.0122	(0.0113)	0.041***	(0.003)
Dislike count	-0.0166**	(0.0072)	-0.017***	(0.001)
Comment count	-0.0108	(0.0079)	-0.006***	(0.002)
Title length	-0.0106	(0.0141)	-0.004**	(0.002)
Description length	0.0100	(0.0075)	0.008***	(0.001)
Tenure	0.0056	(0.0061)	0.004	(0.003)
Followers	-0.0006	(0.0084)	-0.005**	(0.002)
Account total views	-0.0065	(0.0083)	0.0002	(0.002)
Instagram followers	0.0002	(0.0043)	-0.002	(0.002)
Avg. diff. in followers	0.0013	(0.0057)	-0.001*	(0.001)
Avg. diff. in total views	0.0081	(0.0056)	0.001*	(0.001)
Avg. diff in Instagram followers	0.0068*	(0.0040)	0.001***	(0.0003)
Format - GRWM	0.0102	(0.0149)	0.005***	(0.001)
Format - haul	0.0045	(0.0105)	0.006***	(0.001)
Format - review	0.0030	(0.0106)	-0.002	(0.001)
Format - routine	0.0115	(0.0108)	-0.003**	(0.002)
Format - tutorial	0.0133	(0.0119)	0.001	(0.002)
Format - vlog	0.0136	(0.0118)	0.001	(0.002)
Verbal sophistication 1gram	-0.0488***	(0.0168)	-0.013***	(0.003)
Verbal sophistication 2gram	0.0025	(0.0280)	0.024***	(0.005)
Verbal sophistication 3gram	0.0418**	(0.0192)	-0.002	(0.003)
Verbal concreteness	0.1647	(0.1153)	-0.024	(0.026)
Verbal topic 0	-0.0159	(0.0172)	0.002	(0.002)
Verbal topic 1	-0.0097	(0.0397)	0.013***	(0.004)
Verbal topic 2	0.0676**	(0.0329)	0.010**	(0.004)
Verbal topic 3	0.0533	(0.0771)	0.013***	(0.004)
Verbal topic 4	-0.0109	(0.0214)	-0.013***	(0.004)
Verbal topic 5	-0.0010	(0.0494)	-0.001	(0.004)
Verbal topic 6	0.0430	(0.0514)	0.009*	(0.005)
Verbal topic 7	0.0492	(0.0432)	-0.013***	(0.005)
Verbal topic 8	-0.0145	(0.0274)	0.002	(0.005)
Verbal topic 9	0.0232	(0.0489)	-0.023***	(0.005)
Age	-0.0358	(0.0567)	-0.002	(0.005)
Emotion - angry	-0.0563	(0.0399)	-0.002	(0.003)
Emotion - disgust	-0.1190*	(0.0678)	0.008	(0.005)
Emotion - fear	-0.0030	(0.0311)	0.001	(0.002)
Emotion - happy	-0.0129	(0.0176)	0.004***	(0.001)
Emotion - sad	-0.0157	(0.0228)	0.0004	(0.002)

Emotion - surprise	-0.0014	(0.0310)	-0.0002	(0.003)
Race - asian	0.0073	(0.0423)	0.001	(0.004)
Race - black	-0.0206	(0.0530)	-0.002	(0.004)
Race - indian	0.1167	(0.0850)	-0.0002	(0.008)
Race - latino_hispanic	-0.0515	(0.0640)	-0.001	(0.006)
Race - middle_eastern	-0.0152	(0.0557)	0.004	(0.004)
Facial attractiveness	0.0319	(0.0636)	0.0005	(0.006)
Makeup heaviness	0.1958	(0.1855)	0.003	(0.016)
Object – top/t-shirt/sweatshirt	0.0662	(0.0488)	-0.002	(0.004)
Object – jacket			-0.084***	(0.024)
Object – pants	-0.0823	(0.0784)	0.004	(0.006)
Object – skirt			0.050	(0.060)
Object – dress	-0.0156	(0.0135)	0.002**	(0.001)
Object - glasses	0.0181	(0.0555)	-0.0004	(0.006)
Object – hat			0.026	(0.036)
Object – watch			0.016	(0.029)
Object – belt			0.024	(0.036)
Object – tights/stockings	-0.1743	(0.1712)	0.018	(0.016)
Object – shoe	0.0287	(0.0186)	0.002	(0.001)
Object – bag/wallet			-0.007	(0.020)
Object – collar	0.0155	(0.0684)	-0.027	(0.023)
Object – lapel	0.1900**	(0.0964)	0.022	(0.014)
Object – sleeve	-0.0020	(0.0120)	-0.001	(0.001)
Object – pocket			0.029**	(0.014)
Object – neckline	0.0060	(0.0131)	0.001	(0.001)
Foundation1_G	0.0449	(0.0890)	0.004	(0.005)
Foundation2_G	-0.2759***	(0.0938)	0.003	(0.007)
Foundation3_G	0.0372	(0.0777)	0.003	(0.006)
Blush1_G	-0.0202	(0.1174)	-0.005	(0.010)
Blush2_G	0.1448	(0.1582)	-0.002	(0.014)
Blush3_G	-0.1618	(0.2053)	-0.020	(0.019)
Blush4_G	0.4582**	(0.1779)	0.020	(0.023)
Blush5_G	-0.1682	(0.1784)	0.010	(0.016)
Lip_G	0.0327	(0.0544)	0.002	(0.004)
Lipliner_G	-0.0471	(0.0801)	-0.008	(0.006)
Eyeshadow1_G	0.0366	(0.0591)	-0.002	(0.005)
Eyeshadow2_G	-0.1176	(0.1386)	-0.002	(0.010)
Eyeshadow3_G	0.3328	(0.2037)	-0.004	(0.014)
Eyeshadow4_G	-0.3822	(0.2382)	-0.017	(0.016)

Eyeshadow5_G	0.1056	(0.1449)	0.018	(0.012)
Eyeliner_G	-0.1003*	(0.0588)	-0.011**	(0.005)
Eyebrow_G	0.0271	(0.0538)	0.006	(0.005)
NIMA score average	-0.0117	(0.0199)	-0.001	(0.002)
Monthly ad. spending	0.0016	(0.0017)	0.0003*	(0.0002)
sub_diff	-0.0491	(0.0495)		
sub_diff2	0.0301	(0.0197)		
sub_diff3	-0.0026	(0.0016)		
view_diff	0.0208	(0.0249)		
view_diff2	-0.0057	(0.0066)		
view_diff3	0.0003	(0.0004)		
Constant	1.1847***	(0.3010)		
Observations	2,514 (589)		103,479 (1,079)	
Fixed Effects	Individual, year-month (time), 30 brands, 30 product types			
R^2	0.5637		0.3233	
Note:*p<0.1;**p<0.05;***p<0.01				

OA 5.F. FULL RESULTS OF TABLE 11

Table S35 provides all coefficients for the estimation of Equation (10). (The abbreviated results appear in Table 11 in the main paper.)

Table S35. Predictors of a Significant Change in Average Loudness in Sponsored Videos

	DV: Average Loudness	
Variable	Coef.	SE
Sponsorship	**-0.059*** **	**(0.018)**
Sponsorship & Interaction Variables between Sponsorship and Content Variables		
Sponsorship * 1(age <= median)	**-0.001**	**(0.017)**
Sponsorship * 1(tenure <= median)	**-0.003**	**(0.018)**
Sponsorship * 1(N. of followers <= median)	**-0.049** **	**(0.020)**
Sponsorship * 1(N. of total views <= median)	**-0.0004**	**(0.018)**
Sponsorship * 1(N. of Instagram followers <= median)	**0.032**	**(0.020)**
Duration	-0.274***	(0.016)
View count	-0.014	(0.013)
Like count	-0.020	(0.018)
Dislike count	0.015	(0.010)
Comment count	0.050***	(0.010)

Title length	-0.015	(0.016)
Description length	0.068***	(0.016)
Tenure	-0.057	(0.039)
Followers	-0.055**	(0.025)
Total views	-0.011	(0.023)
Instagram followers	0.026	(0.018)
Avg. diff. in followers	-0.004	(0.006)
Avg. diff. in total views	-0.001	(0.008)
Avg. diff in Instagram followers	-0.0004	(0.005)
Format - GRWM	0.044***	(0.015)
Format - haul	0.073***	(0.013)
Format - review	-0.047***	(0.013)
Format - routine	-0.069***	(0.012)
Format - tutorial	-0.037***	(0.014)
Format - vlog	0.098***	(0.015)
Verbal sophistication 1gram	0.160***	(0.040)
Verbal sophistication 2gram	-0.218***	(0.056)
Verbal sophistication 3gram	-0.012	(0.029)
Verbal concreteness	-0.594***	(0.196)
Verbal topic 0	-0.084***	(0.020)
Verbal topic 1	0.060	(0.039)
Verbal topic 2	-0.102***	(0.037)
Verbal topic 3	-0.050	(0.031)
Verbal topic 4	-0.098***	(0.026)
Verbal topic 5	-0.110***	(0.029)
Verbal topic 6	0.001	(0.030)
Verbal topic 7	-0.199***	(0.040)
Verbal topic 8	0.016	(0.035)
Verbal topic 9	0.226	(0.213)
Age	-0.033	(0.031)
Emotion - angry	0.009	(0.048)
Emotion - disgust	0.035	(0.026)
Emotion - fear	0.001	(0.016)
Emotion - happy	0.033	(0.022)
Emotion - sad	-0.021	(0.026)
Emotion - surprise	0.021	(0.038)
Race – asian	0.005	(0.054)
Race - black	0.041	(0.069)
Race - indian	-0.059	(0.061)

Race - latino_hispanic	0.074	(0.051)
Race - middle_eastern	0.080	(0.060)
Facial attractiveness	0.009	(0.231)
Makeup heaviness	0.226	(0.213)
Object – top/t-shirt/sweatshirt	0.077	(0.055)
Object – jacket		
Object – pants	0.050	(0.100)
Object – skirt	-0.087	(0.184)
Object – dress	0.030**	(0.013)
Object - glasses	-0.076	(0.054)
Object – hat	-0.359	(0.268)
Object – watch	-0.167	(0.150)
Object – belt	-0.132	(0.192)
Object – tights/stockings	0.417	(0.355)
Object – shoe	0.050***	(0.019)
Object – bag/wallet	-0.040	(0.135)
Object – collar	0.894*	(0.512)
Object – lapel	-0.304**	(0.138)
Object – sleeve	0.025**	(0.011)
Object – pocket	-0.199	(0.240)
Object – neckline	-0.011	(0.011)
Foundation1_G	0.041	(0.070)
Foundation2_G	-0.022	(0.074)
Foundation3_G	-0.047	(0.065)
Blush1_G	0.124	(0.120)
Blush2_G	0.250	(0.179)
Blush3_G	-0.243	(0.214)
Blush4_G	0.171	(0.228)
Blush5_G	-0.487***	(0.165)
Lip_G	0.002	(0.050)
Lipliner_G	0.043	(0.080)
Eyeshadow1_G	-0.068	(0.053)
Eyeshadow2_G	-0.241**	(0.100)
Eyeshadow3_G	0.333**	(0.150)
Eyeshadow4_G	-0.182	(0.187)
Eyeshadow5_G	-0.126	(0.131)
Eyeliner_G	0.088	(0.056)
Eyebrow_G	0.151***	(0.047)
NIMA score average	0.368***	(0.027)

Monthly ad. spending	0.001	(0.002)
Constant	1.538*	(0.880)
Observations (Influencer)	32,920 (1,017)	
Fixed Effects	Individual, year-month (time), 30 brands, 30 product types	
R^2	0.3462	
Note:*p<0.1;**p<0.05;***p<0.01		

OA SECTION 6. ROBUSTNESS AND VALIDITY CHECKS

OA 6.A. IV MODEL WITH ADDITIONAL INSTRUMENTS FOR VOICES

In light of the two possible sources of endogeneity: the first originating from the sponsorship variable, and the second emanating from vocal characteristics, we consider a total of 11 endogenous regressors, which include five vocal characteristics, sponsorship, and five interaction variables between sponsorship and the vocal characteristics. We employ five key variables that contribute to weather as the instruments for the vocal characteristics designated by NOAA Global Historical Climatology Network Daily (GHCN-D; Menne et al., 2012): the minimum and maximum of the temperature, and precipitation. To address the presence of these endogenous variables, we utilize a set of 11 instruments, comprising two instruments for sponsorship (namely, the weekly number of sponsorships for the brand and parent company, respectively), three instruments for voice (namely, max temperature, min temperature, and precipitation), and six interactions between two instruments for sponsorship and three instruments for voice. In Table S36, Our results indicate that the coefficient of the instrumented interaction variable between sponsorship and the average loudness is significantly negative, suggesting that reducing the average loudness in sponsored videos can lead to an increase in consumer sentiment. These findings lend further support to the notion that vocal characteristics and their interaction with sponsorship play an essential role in shaping consumer sentiment towards sponsored content.

Table S36. Effects of Sponsorship, Vocal Characteristics, and Interactions on Sentiment: IV Model with Multiple Regressors and IVs

Vocal characteristics	Coef	SE
Average loudness	0.8096**	(0.3275)
Average pitch	-1.5780**	(0.7871)
Loudness variability	-0.5082***	(0.1920)
Pitch variability	1.5381**	(0.7647)
Talking duration	1.5997*	(0.8385)
Sponsorship & Sponsored Vocal Characteristics		
Sponsorship	0.1149**	(0.0474)
*Sponsorship * Average loudness*	-0.2927**	(0.1221)
*Sponsorship * Average pitch*	1.1527*	(0.6971)
*Sponsorship * Loudness variability*	0.1594	(0.0979)
*Sponsorship * Pitch variability*	-1.1467	(0.7398)
*Sponsorship * Talking duration*	-2.0387**	(0.9779)
Constant	-0.1975	(0.9175)
Individual/Time/Content FEs (visual, verbal, textual, and popularity)	Yes	
Observations (Individual)	32,920 (1,017)	
R^2	0.3351	

OA 6.B. IV MODEL WITH INTERACTIONS BETWEEN SPONSORSHIP AND VOCAL AND VISUAL CONTENT VARIABLES

In Table S37, we extended our analysis by incorporating interaction variables between sponsorship and vocal/visual characteristics, as well as other influencer-related content variables. This approach allowed us to thoroughly examine the effects of sponsorship on vocal characteristics and their potential interactions with other relevant factors, thereby enhancing the validity and reliability of our results.

Table S37. Effects of Sponsorship, Vocal Characteristics, and Interactions on Sentiment: IV Model with Sponsorship-Content Variable Interactions

Variable	Coef.	SE
Average loudness	**0.0023***	**(0.0013)**
Average pitch	**0.0063***	**(0.0035)**
Loudness variability	**-0.0026**	**(0.0016)**
Pitch variability	**-0.0031**	**(0.0029)**
Talking duration	**-0.0092*****	**(0.0034)**
Sponsorship & Sponsored Vocal Characteristics		
Sponsorship	**-0.1722**	**(0.1107)**
Sponsorship * Average loudness	**-0.0049****	**(0.0024)**
Sponsorship * Average pitch	**-0.0080**	**(0.0061)**
Sponsorship * Loudness variability	**-0.0022**	**(0.0029)**
Sponsorship * Pitch variability	**0.0025**	**(0.0051)**
Sponsorship * Talking duration	**0.0060**	**(0.0059)**
Other Content Control Variables		
Duration	0.0011	(0.0026)
View count	-0.0235***	(0.0039)
Like count	0.0296***	(0.0047)
Dislike count	-0.0145***	(0.0024)
Comment count	-0.0080***	(0.0026)
Title length	-0.0106***	(0.0039)
Description length	0.0085***	(0.0021)
Tenure	0.0027	(0.0041)
Number of followers	-0.0070**	(0.0031)
Number of total views	0.0020	(0.0032)
Number of Instagram followers	-0.0007	(0.0021)
Avg. diff. in the number of followers	0.0016	(0.0017)
Avg. diff. in the number of total views	0.0005	(0.0017)

Avg. diff. in the number of Instagram followers	0.0019	(0.0014)
Format - GRWM	0.0069***	(0.0022)
Format - haul	0.0059***	(0.0019)
Format - review	-0.0017	(0.0020)
Format - routine	-0.0047**	(0.0020)
Format - tutorial	-0.0015	(0.0019)
Format - vlog	0.0006	(0.0021)
Verbal sophistication 1gram	-0.0077***	(0.0028)
Verbal sophistication 2gram	0.0069	(0.0048)
Verbal sophistication 3gram	0.0035	(0.0034)
Verbal concreteness	0.0093	(0.0180)
Verbal topic 0	-0.0002	(0.0029)
Verbal topic 1	0.0034	(0.0064)
Verbal topic 2	0.0107	(0.0081)
Verbal topic 3	0.0158***	(0.0054)
Verbal topic 4	-0.0166***	(0.0060)
Verbal topic 5	-0.0142*	(0.0075)
Verbal topic 6	0.0040	(0.0093)
Verbal topic 7	-0.0204**	(0.0085)
Verbal topic 8	0.0182**	(0.0091)
Verbal topic 9	-0.0162*	(0.0097)
Age	-0.0179	(0.0133)
Emotion - angry	-0.0272***	(0.0090)
Emotion - disgust	0.0002	(0.0172)
Emotion - fear	-0.0016	(0.0077)
Emotion - happy	0.0019	(0.0045)
Emotion - sad	-0.0026	(0.0059)
Emotion - surprise	-0.0047	(0.0088)
Race - asian	0.0168**	(0.0082)
Race - black	0.0131	(0.0089)
Race - indian	0.0315*	(0.0169)
Race - latino_hispanic	0.0211	(0.0149)
Race - middle_eastern	0.0279**	(0.0126)
Facial attractiveness	0.0218	(0.0165)
Makeup heaviness	-0.0977*	(0.0516)
Object – top/t-shirt/sweatshirt	0.0166	(0.0104)
Object – jacket		
Object – pants	0.0218	(0.0210)
Object – skirt	-0.0126	(0.0724)
Object – dress	0.0048	(0.0037)

Object - glasses	0.0103	(0.0190)
Object – hat	0.0580	(0.0894)
Object – watch	-0.0294	(0.1248)
Object – belt	0.0323	(0.0719)
Object – tights/stockings	-0.0456	(0.0712)
Object – shoe	-0.0003	(0.0048)
Object – bag/wallet	-0.0164	(0.0572)
Object – collar	-0.0236	(0.0489)
Object – lapel	-0.1182	(0.1808)
Object – sleeve	0.0010	(0.0032)
Object – pocket	0.0467	(0.0427)
Object – neckline	0.0001	(0.0033)
Foundation1 (G)	-0.0032	(0.0172)
Foundation2 (G)	0.0074	(0.0196)
Foundation3 (G)	0.0081	(0.0170)
Blush1 (G)	-0.0105	(0.0274)
Blush2 (G)	0.0342	(0.0434)
Blush3 (G)	-0.1093*	(0.0609)
Blush4 (G)	0.0812	(0.0701)
Blush5 (G)	-0.0105	(0.0463)
Lip (G)	-0.0105	(0.0140)
Lipliner (G)	0.0260	(0.0214)
Eyeshadow1 (G)	-0.0119	(0.0147)
Eyeshadow2 (G)	0.0645**	(0.0322)
Eyeshadow3 (G)	-0.1029**	(0.0505)
Eyeshadow4 (G)	0.0647	(0.0610)
Eyeshadow5 (G)	-0.0196	(0.0403)
Eyeliner (G)	0.0042	(0.0156)
Eyebrow (G)	-0.0075	(0.0141)
NIMA score average	-0.0028	(0.0046)
Monthly ad. spending	0.0001	(0.0003)
Interactions Between Sponsorship and Other Content Control Variables		
Sponsorship*Emotion - angry	0.0446***	(0.0154)
Sponsorship*Emotion - disgust	0.0228	(0.0286)
Sponsorship*Emotion - fear	0.0045	(0.0133)
Sponsorship*Emotion - happy	0.0098	(0.0076)
Sponsorship*Emotion - sad	0.0102	(0.0102)
Sponsorship*Emotion - surprise	0.0060	(0.0150)
Sponsorship*Race - asian	-0.0076	(0.0115)
Sponsorship*Race - black	-0.0094	(0.0127)

Sponsorship*Race - indian	-0.0271	(0.0264)
Sponsorship*Race - latino_hispanic	-0.0202	(0.0237)
Sponsorship*Race - middle_eastern	-0.0504**	(0.0201)
Sponsorship*Makeup heaviness	0.1744**	(0.0888)
Sponsorship*Object – top/t-shirt/sweatshirt	-0.0199	(0.0172)
Sponsorship*Object – jacket		
Sponsorship*Object – pants	-0.0288	(0.0328)
Sponsorship*Object – skirt	0.0425	(0.1435)
Sponsorship*Object – dress	-0.0078	(0.0061)
Sponsorship*Object - glasses	-0.0543*	(0.0307)
Sponsorship*Object – hat	-0.0054	(0.1156)
Sponsorship*Object – watch		
Sponsorship*Object – belt		
Sponsorship*Object – tights/stockings	0.0418	(0.1145)
Sponsorship*Object – shoe	0.0116	(0.0084)
Sponsorship*Object – bag/wallet	-0.0672	(0.1276)
Sponsorship*Object – collar	0.0787	(0.0996)
Sponsorship*Object – lapel	0.1899	(0.2261)
Sponsorship*Object – sleeve	-0.0058	(0.0054)
Sponsorship*Object – pocket	-0.0156	(0.0621)
Sponsorship*Object – neckline	0.0001	(0.0057)
Sponsorship*Blush1 (G)	0.0440	(0.0484)
Sponsorship*Blush2 (G)	-0.0644	(0.0767)
Sponsorship*Blush3 (G)	0.1632	(0.1053)
Sponsorship*Blush4 (G)	-0.0632	(0.1239)
Sponsorship*Blush5 (G)	-0.0348	(0.0812)
Sponsorship*Eyebrow (G)	0.0169	(0.0237)
Sponsorship*Eyeliner (G)	-0.0373	(0.0271)
Sponsorship*Eyeshadow1 (G)	0.0111	(0.0246)
Sponsorship*Eyeshadow2 (G)	-0.0821	(0.0573)
Sponsorship*Eyeshadow3 (G)	0.2035**	(0.0885)
Sponsorship*Eyeshadow4 (G)	-0.2401**	(0.1088)
Sponsorship*Eyeshadow5 (G)	0.1173	(0.0716)
Sponsorship*Foundation1 (G)	-0.0267	(0.0299)
Sponsorship*Foundation2 (G)	-0.0058	(0.0335)
Sponsorship*Foundation3 (G)	0.0112	(0.0293)
Sponsorship*Lip (G)	0.0129	(0.0234)
Sponsorship*Lipliner (G)	-0.0513	(0.0364)
Sponsorship*Avg. diff. in the number of followers	-0.0058**	(0.0027)
Sponsorship*Avg. diff. in the number of Instagram followers	-0.0036	(0.0025)

Sponsorship*Avg. diff. in the number of total views	-0.0014	(0.0028)
Sponsorship*Age	0.0446**	(0.0219)
Sponsorship*Facial attractiveness	-0.0519*	(0.0279)
Sponsorship*Comment count	0.0020	(0.0043)
Sponsorship*Description length	0.0047	(0.0038)
Sponsorship*Dislike count	-0.0055	(0.0038)
Sponsorship*Duration	0.0027	(0.0044)
Sponsorship*Like count	0.0046	(0.0067)
Sponsorship*Tenure	-0.0006	(0.0046)
Sponsorship*Title length	0.0143**	(0.0066)
Sponsorship*View count	-0.0012	(0.0055)
Sponsorship*NIMA score average	0.0012	(0.0075)
Sponsorship*Number of followers	0.0081**	(0.0036)
Sponsorship*Number of Instagram followers	-0.0035**	(0.0015)
Sponsorship*Number of total views	-0.0049**	(0.0024)
Constant	0.8299***	(0.0952)
Observations (Influencer)	32,920 (1,017)	
Fixed Effects	Individual, time (year-month) 30 brands, 30 product types	
R^2	0.3376	

OA 6.C. ALTERNATIVE DV – NUMBER OF SPONSORSHIP DEALS

In the main analyses, we used consumer sentiment as the outcome variable because it is predictive of brand performance metrics such as sales conversion (Schneider and Gupta 2016) and is widely used in marketing research. Now, we implement an OLS model to test the impact of voice on the outcome variable of interest, *the logarithm of the number of future sponsorship deals from brand* m.

In Equation (S13), we instrument sponsorship and the interaction between sponsorship and vocal characteristics. Then, we use the instrumented sponsorship and interaction to estimate Equation (S14), in which the dependent variable, $Log\ of\ Number\ of\ sponsorship\ deals\ (brand\ m)_{ijt}$. $Log\ of\ Number\ of\ sponsorship\ deals(brand\ m)_{ijt}$ is the logarithm of the number of sponsorship deals that influencer i received from brand m after posting video j at time period t.

$$Sponsorship_{ijt} = \beta_0 + \beta_1 \times Number\ of\ videos(brand\ m) + \beta_2 \times \overrightarrow{Voice_{ijt}^{k}} \times Number\ of\ videos(brand\ m) + \vec{\Gamma} \times \overrightarrow{ControlVariables} + I_i + \tau_t + \epsilon_{ijt}$$

$$Sponsorship_{ijt} \times \overrightarrow{Voice_{ijt}^{k}} = \beta_0 + \beta_1 \times Number\ of\ videos(brand\ m) + \beta_2 \times \overrightarrow{Voice_{ijt}^{k}} \times Number\ of\ videos(brand\ m) + \vec{\Gamma} \times \overrightarrow{ControlVariables} + I_i + \tau_t + \epsilon_{ijt} \tag{S13}$$

$$Log\ of\ Number\ of\ sponsorship\ deals(brand\ m)_{ijt} = \beta_0 + \beta_1 \times \widetilde{Sponsorship}_{ijt} + \beta_2 \times \overrightarrow{Voice_{ijt}^{k}} + \beta_3 \times \widetilde{Sponsorship}_{ijt} \times \overrightarrow{Voice_{ijt}^{k}} + \vec{\Gamma} \times \overrightarrow{ControlVariables} + I_i + \tau_t + \epsilon_{ijt} \tag{S14}$$

Table S38 shows the results of the model with the log of the number of sponsorship deals from the same brand as our DV. The coefficients of average loudness are 0.0047 for non-sponsored videos and -0.0137 (= 0.0047 - 0.0184) for sponsored videos. The findings imply that an increase in the average loudness in a sponsored video hurts the influencer's chance of getting another sponsorship from the same brand in the future. This is consistent with our findings on the relationship between consumer sentiment and average loudness in sponsored videos.

Table S38. Effect of Vocal Characteristics in Sponsored Videos on Future Sponsorship Deals

	DV: Log of Number of Sponsorship Deals from *brand m*	
Vocal Characteristics	Coef.	SE
Average loudness	**0.0047**	**(0.0047)**
Average pitch	**0.0278****	**(0.0129)**
Loudness variability	**-0.0254*****	**(0.0059)**
Pitch variability	**-0.0093**	**(0.0108)**
Talking duration	**-0.0280****	**(0.0125)**
Sponsorship & Sponsored Vocal Characteristics		
Sponsorship	**0.4781*****	**(0.0096)**
Sponsorship * Average loudness	**-0.0184****	**(0.0085)**

Sponsorship * Average pitch	**-0.0456****	**(0.0224)**
Sponsorship * Loudness variability	**0.0410*****	**(0.0105)**
Sponsorship * Pitch variability	**0.0123**	**(0.0190)**
Sponsorship * Talking duration	**0.0140**	**(0.0216)**
Constant	-1.0834***	(0.2916)
Observations (Influencer)	32,920 (1,017)	
Fixed Effects	Influencer, year-month, and other content control variables used in OA Section 5	
R^2	0.3133	
Notes: Table entries are coefficients, and the robust standard errors are in parentheses. *p < 0.1; **p < 0.05; ***p < 0.01.		

OA 6.D. VOCAL IMPACTS OF TWO INFLUENCER GROUPS BY SPONSORSHIP COEFFICIENT SIGNS (+/-)

We further investigate how increasing (or decreasing) their loudness in sponsored videos hurts (or mitigates) consumer sentiment by creating two influencer subsamples through the influencer fixed effects I_i and the sponsorship coefficient, β_1, in Equation (S15). Specifically, by summing up the two coefficients, I_i and β_1, we construct two influencer subsamples: (1) influencers with negative coefficients ($Influencers\_with\_negative\_coefs$) and (2) influencers with positive coefficients ($Influencers\_\ with\_positive\_coefs$).

$$Loudness_{ijt} = \beta_0 + \beta_1 \times Sponsorship_{ijt} + \vec{\Gamma} \times \overrightarrow{ControlVariables}_{ijt} + I_i + \tau_t + \epsilon_{ijt} \text{ for all } i \in \{1, \dots, I\}$$
$$\begin{cases} \beta_1 + I_i < 0, & then\ Influencers\_with\_negative\_coefs \\ \beta_1 + I_i \geq 0, & then\ Influencers\_with\_positive\_coefs \end{cases} \tag{S15}$$

In Equation (S16), $Influencers\_with\_negative\_coefs$ are influencers whose sum of sponsorship coefficient and individual fixed effects was negative ($\beta_1 + I_i < 0$), and $Influencers\_with\_positive\_coefs$ are those whose sum of the two coefficients was positive ($\beta_1 + I_i \geq 0$). We use the IV model in Equation (S16) to estimate the sponsorship coefficients for each group, and we report the results in Table S39.

$$\begin{aligned} Sentiment_{ijt}^{Type} &= \beta_0 + \beta_1 \times \widetilde{Sponsorship}_{ijt}^{Type} + \beta_2 \times \overrightarrow{Voice_{ijt}^{k}}^{Type} \\ &+ \beta_3 \times \widetilde{Sponsorship}_{ijt} \times \overrightarrow{Voice_{ijt}^{k}}^{Type} + \vec{\Gamma} \times \overrightarrow{ControlVariables} \\ &+ I_i + \tau_t + \epsilon_{ijt} \end{aligned} \tag{S16}$$

where $Type \in \{Influencers\_with\_negative\_coefs, Influencers\_with\_positive\_coefs\}$

We find that influencers who increased their loudness in sponsorship videos are hurt more from sponsorship disclosure. Specifically, the influencers who increase their loudness in sponsored videos ($\beta_1 + I_i \geq 0$) suffer from a decrease of 0.016 in consumer sentiment, while the influencers who reduce their loudness in sponsored videos ($\beta_1 + I_i < 0$) experience a decrease of only 0.012. Moreover, for the influencers who increase their loudness in sponsored videos, one unit increase in loudness is likely to hurt by 0.005 on their consumer sentiment. This robustness check again confirms one of our main results that increasing the loudness hurts consumer sentiment in the sponsored videos.

Table S39. Sponsorship & Loudness Effect on Consumer Sentiment in Two Influencer Groups

	(1) Influencers whose loudness decrease in sponsored videos		(2) Influencers whose loudness increase in sponsored videos	
Vocal Characteristics	Coef.	SE	Coef.	SE
Average loudness	0.002**	(0.001)	0.002*	(0.001)
Average pitch	0.003*	(0.002)	0.0005	(0.002)
Loudness variability	-0.002***	(0.001)	-0.0001	(0.001)
Pitch variability	-0.001	(0.002)	0.002	(0.002)
Talking duration	-0.006***	(0.002)	-0.003	(0.002)
Sponsorship & Sponsored Vocal Characteristics				
Sponsorship	-0.012***	(0.002)	-0.016***	(0.002)
Sponsorship*Average loudness	-0.001	(0.002)	-0.005**	(0.002)
Sponsorship*Average pitch	-0.005	(0.005)	-0.001	(0.005)
Sponsorship*Loudness variability	-0.001	(0.002)	-0.005*	(0.003)
Sponsorship*Pitch variability	0.001	(0.004)	-0.003	(0.004)
Sponsorship*Talking duration	0.003	(0.005)	0.001	(0.005)
Constant	0.765***	(0.056)	0.692***	(0.056)
Observations (Influencer)	56,242 (514)		44,921 (480)	
Fixed Effects	Influencer, year-month, and other content control variables used in OA Section 5			
R^2	0.3433		0.3201	
Notes: The first group contains influencers whose sum of sponsorship and influencer-level fixed effect coefficients is negative; the second group contains influencers whose sum of sponsorship and influencer-level fixed effect coefficients is positive. *p < 0.1; **p < 0.05; ***p < 0.01				

OA 6.E. VOCAL FEATURE COMPARISONS WITH REVERSE CAUSALITY CHECK

This section addresses the potential issue of reverse causality in the first-stage regression analysis. While we have assumed that influencers strategically modulate their vocal characteristics in response to sponsorship, it is plausible to consider whether brands intentionally sponsor influencers who exhibit adeptness in manipulative vocal strategies.[10] To examine this possibility, we conducted a comparison of vocal characteristics in non-sponsored videos between the treatment group (i.e., influencers who received sponsorship) and the control group (i.e., influencers who did not receive sponsorship). If there is no discernible difference in vocal characteristics between the organic (non-sponsored) videos of the two influencer groups in the absence of treatment (i.e., brand sponsorship), it would suggest that firms do not use vocal characteristics as a criterion for screening influencers and making sponsorship decisions, thereby negating the possibility of reverse causality. The results of the comparison are presented in Table S40, where none of the t-statistics are found to be significant, leading us to fail to reject the null hypothesis that there is no difference in vocal attributes between the non-sponsored videos of the treatment and control groups. Consequently, we can rule out the possibility of reverse causality in our findings.

Table S40. Two Sample T-test of Normalized Vocal Characteristics

Control group's non-sponsored video vs. Treatment group's non-sponsored video	Loudness	Pitch	Loudness Variability	Pitch Variability	Talking Duration
T-statistics (p-value)	-0.43905	-0.096211	0.070527	0.36778	-0.15251
	(0.6606)	(0.9234)	(0.9438)	(0.713)	(0.8788)
95% confidence interval	(-0.0107,	(-0.0094,	(-0.0084,	(-0.0073,	(-0.0096,
	0.0068)	0.0085)	0.009)	0.0106)	0.0082)
Notes. The table entries without parentheses are t-statistics, and the entries in parentheses are the corresponding p-value. *p < 0.10; **p < 0.05; ***p < 0.01.					

OA 6.F. RD VALIDITY CHECK: NO MANIPULATION OF CONTENT CONTROLS

One concern about the validity of FRD is that influencers can fully observe the assignment variables and change their content characteristics discontinuously around the assignment variables' cutoffs to manipulate their own sponsorship status. For instance, influencers who have exceeded the follower count threshold (50,000) might become more relaxed about creating content and change their content characteristics, while influencers just below the threshold work hard on content creation because they want to get above the threshold. We test for the possibility of manipulation by checking for a relationship between the influencer's sponsorship status and content control characteristics (Mukherjee et al. 2018).

[10] Later, through our interviews with Neoreach, we learned that firms decide whether to sponsor an influencer based on the influencer's popularity and other metrics, not based on whether the influencer modulates their voice in sponsored videos. However, to show that the reverse causality does not hold empirically, we assume in this robustness test that we do not know *a priori* about how firms choose influencers for sponsorships.

$$Content\ characteristics_{ijt} = \beta_0 + \beta_1 * Sponsorship_{ijt} + \vec{\Gamma} \times \overrightarrow{OtherControlVariables} + I_i + \tau_t + \epsilon_{ijt} \quad \text{(S18)}$$

We estimate Equation (S18) on two visual and verbal characteristics: the logarithm of the spoken word frequency via COCA corpus, and the neural image assessment (NIMA) score. In Table S41, we find no effect of sponsorship on the content characteristics.

Table S41. FRD Validity Check: Effect of Sponsorship on Content Characteristics

	FRD Model			
	Visual Content (1)		Verbal Content (2)	
Sponsorship	-0.036	(0.026)	-0.256	(0.529)
Constant	-0.231	(0.614)	-2.118	(6.276)
Observations (Influencer)	2,514 (589)		2,514 (589)	
Fixed Effects	Influencer, year-month, and other content control variables used in OA Section 5			
R^2	0.989		0.182	

OA 6.G. RD VALIDITY CHECK: PLACEBO THRESHOLDS

In addition to the anticipated discontinuities in a regression discontinuity (RD) model, there may exist other sources of discontinuity. One such source is the potential influence of consumers' observations of an influencer's followership on their evaluations of the influencer and their sentiment toward the influencer's videos. If consumers rely on the number of followers as a heuristic for assessing an influencer, then the RD model may encounter an unintended discontinuity. Thus, it is important to account for potential sources of discontinuity when analyzing the effectiveness of an RD model.

We address the possibility of other discontinuities by implementing placebo thresholds, as proposed by Eggers et al. (2018). Specifically, we calculate the midpoint between the actual threshold and the maximum value for each brand and assignment variable (i.e., the number of followers and average views). To illustrate, for the brand Vanity Planet, the sponsorship thresholds are 7,114 followers and 995 average views, while the maximum values are 3,974,312 followers and 13,569,346 average views. Therefore, we obtain midpoint values of 1,990,713 followers and 6,785,171 views. This approach allows us to test whether our estimated effects at the actual threshold are significantly different from the effects at these placebo thresholds.

To examine the impact of exceeding the placebo thresholds on consumer sentiment, we employ ten brands in our FRD models. Specifically, we investigate whether there are changes in consumer sentiment as a result of exceeding the placebo thresholds.

$$\begin{aligned} Consumer\ sentiment_{ijt} &= \beta_0 + \beta_1 * 1\left(X_{it}^f > placebo_{it}^f\right) + \beta_2 * 1(X_{it}^v > placebo_{it}^v) \\ &+ \vec{\Gamma} \times \overrightarrow{ControlVariables} + I_i + \tau_t + \epsilon_{ijt} \end{aligned} \quad \text{(S19)}$$

where $1\left(X_{it}^f > placebo_{it}^f\right)$ and $1(X_{it}^v > placebo_{it}^v)$ indicate whether the follower count and average views of influencer *i* exceed the placebo threshold at time *t*. In Table S42, we show that neither placebo threshold causes a significant change in consumer sentiment.

Table S42. Effect of Placebo Thresholds on Consumer Sentiment

	DV: Consumer Sentiment	
Variable	Coef.	SE
1(followers > Placebo followers)	0.006	(0.033)
1(views > Placebo views)	0.012	(0.029)
Constant	1.395***	(0.373)
Observations (Individual)	2,514 (589)	
Fixed Effects	Influencer, year-month, and other content control variables used in OA Section 5	
R^2	0.5619	
Notes. *p < 0.1; **p < 0.05; ***p < 0.01.		

We also test the Sharp RD model with a placebo threshold of 300,000 followers (the median follower count of the sponsored influencers) and the corresponding minimum average views (995) for the brand Vanity Planet. The estimated coefficient of sponsorship x average loudness is not significant with the placebo threshold (Table S43).

Table S43. Effect of Placebo Thresholds on the Voice x Sponsorship Interaction Effect on Sentiment

	DV: Consumer Sentiment	
Vocal characteristics	β	SE
Average loudness	-0.014	(0.027)
Average pitch	0.055	(0.143)
Loudness variability	0.005	(0.028)
Pitch variability	-0.080	(0.093)
Talking duration	-0.064	(0.154)
Sponsorship & Sponsored Vocal Characteristics		
Sponsorship	-0.017	(0.039)

*Sponsorship * Average loudness*	0.097	(0.048)
*Sponsorship * Average pitch*	-0.283	(0.182)
*Sponsorship * Loudness variability*	-0.034	(0.030)
*Sponsorship * Pitch variability*	0.219	(0.131)
*Sponsorship * Talking duration*	0.238	(0.146)
Constant	31.427	(58.113)
Individual/Time FEs	Yes	
Content FEs (visual, verbal, textual, popularity)	Yes	
Observations (Individual)	198 (53)	
R^2	0.9816	
Notes: Table entries are coefficients, and robust standard errors are in parentheses. *p < 0.05; **p < 0.01; ***p < 0.001		

OA 6.H. IV MODELS WITH UNMATCHED DATA WITH PROGRESSIVE ADDITIONS OF CONTROL VARIABLES

One may wonder if our main results may change if we used (1) a different set of control variables, and (2) the unmatched data. To address both concerns, we implement the same first-stage and second-stage regressions (Equations 3 & 4 in the main paper) with the progressive addition of control variables using the unmatched data. Table S44 provides the IV results. The unmatched data includes 101,512 observations from 1,017 influencers. The sponsorship coefficient changes from -0.016 to -0.014 with the addition of control variables, but in all columns, a decrease in average loudness improves consumer sentiment in sponsored videos (coefficient = -0.003 to -0.004).

Table S44. Effect of Voice & Sponsorship on Sentiment in the Unmatched Data: IV Model

	DV: Consumer Sentiment							
Variable	(1)		(2)		(3)		(4)	
Vocal Characteristics	β	SE	β	SE	β	SE	β	SE
Average loudness	**0.001****	**(0.001)**	**0.002*****	**(0.001)**	**0.002*****	**(0.001)**	**0.002*****	**(0.001)**
Average pitch	**0.001**	**(0.001)**	**0.001**	**(0.001)**	**0.002**	**(0.001)**	**0.002**	**(0.001)**
Loudness variability	**-0.002*****	**(0.001)**	**-0.001***	**(0.001)**	**-0.001***	**(0.001)**	**-0.001***	**(0.001)**
Pitch variability	**0.002**	**(0.001)**	**0.001**	**(0.001)**	**0.001**	**(0.001)**	**0.001**	**(0.001)**
Talking duration	**-0.004*****	**(0.001)**	**-0.004*****	**(0.001)**	**-0.005*****	**(0.001)**	**-0.005*****	**(0.001)**
Sponsorship & Sponsored Vocal Characteristics								
Sponsorship	**-0.016*****	**(0.001)**	**-0.014*****	**(0.001)**	**-0.014*****	**(0.001)**	**-0.014*****	**(0.001)**
*** Average loudness**	**-0.004****	**(0.002)**	**-0.003****	**(0.001)**	**-0.003****	**(0.001)**	**-0.004****	**(0.001)**
*** Average pitch**	**-0.002**	**(0.004)**	**-0.002**	**(0.004)**	**-0.003**	**(0.004)**	**-0.003**	**(0.004)**
*** Loudness variability**	**-0.003**	**(0.002)**	**-0.003***	**(0.002)**	**-0.003***	**(0.002)**	**-0.003***	**(0.002)**

*** Pitch variability**	**-0.001**	**(0.003)**	**-0.001**	**(0.003)**	**-0.001**	**(0.003)**	**-0.0004**	**(0.003)**
*** Talking duration**	**0.001**	**(0.004)**	**0.001**	**(0.004)**	**0.002**	**(0.004)**	**0.002**	**(0.004)**
Metadata & Ad Spending Control Variables								
Duration			0.006***	(0.001)	0.006***	(0.001)	0.006***	(0.001)
View count			-0.033***	(0.001)	-0.032***	(0.001)	-0.032***	(0.001)
Like count			0.042***	(0.002)	0.041***	(0.002)	0.041***	(0.002)
Dislike count			-0.017***	(0.001)	-0.017***	(0.001)	-0.017***	(0.001)
Comment count			-0.006***	(0.001)	-0.006***	(0.001)	-0.006***	(0.001)
Title length			-0.004***	(0.001)	-0.005***	(0.001)	-0.005***	(0.001)
Description length			0.009***	(0.001)	0.008***	(0.001)	0.008***	(0.001)
Tenure			0.003	(0.002)	0.003	(0.002)	0.003	(0.002)
Number of followers			-0.004***	(0.002)	-0.004***	(0.002)	-0.004***	(0.002)
Number of total views			-0.0003	(0.002)	-0.0003	(0.002)	-0.0003	(0.002)
Number of IG followers			-0.002	(0.001)	-0.002	(0.001)	-0.002	(0.001)
Avg. diff. in followers			-0.002***	(0.001)	-0.002***	(0.001)	-0.002***	(0.001)
Avg. diff. in total views			0.001**	(0.001)	0.001**	(0.001)	0.001**	(0.001)
Avg. diff. in IG followers			0.001***	(0.0003)	0.001***	(0.0003)	0.001***	(0.0003)
Format - GRWM			0.005***	(0.001)	0.005***	(0.001)	0.005***	(0.001)
Format - haul			0.006***	(0.001)	0.005***	(0.001)	0.005***	(0.001)
Format - review			-0.001	(0.001)	-0.001	(0.001)	-0.001	(0.001)
Format - routine			-0.005***	(0.001)	-0.004***	(0.001)	-0.004***	(0.001)
Format - tutorial			0.001	(0.001)	0.001	(0.001)	0.001	(0.001)
Format - vlog			0.001	(0.001)	0.001	(0.001)	0.001	(0.001)
Monthly ad. spending			0.0003**	(0.0002)	0.0003**	(0.0002)	0.0003**	(0.0002)
Verbal Control Variables								
Verbal sophi. 1gram					-0.012***	(0.002)	-0.013***	(0.002)
Verbal sophi. 2gram					0.023***	(0.003)	0.023***	(0.003)
Verbal sophi. 3gram					-0.001	(0.002)	-0.002	(0.002)
Verbal concreteness					-0.024**	(0.012)	-0.023*	(0.012)

Verbal topic 0					0.001	(0.002)	0.001	(0.002)
Verbal topic 1					0.014***	(0.003)	0.013***	(0.003)
Verbal topic 2					0.011***	(0.004)	0.011***	(0.004)
Verbal topic 3					0.013***	(0.003)	0.013***	(0.003)
Verbal topic 4					-0.013***	(0.004)	-0.013***	(0.004)
Verbal topic 5					0.001	(0.004)	0.001	(0.004)
Verbal topic 6					0.008**	(0.004)	0.008**	(0.004)
Verbal topic 7					-0.013***	(0.005)	-0.013***	(0.005)
Verbal topic 8					0.003	(0.005)	0.003	(0.005)
Verbal topic 9					-0.023***	(0.005)	-0.023***	(0.005)
Visual Control Variables								
Age							-0.002	(0.004)
Emotion - angry							-0.002	(0.003)
Emotion - disgust							0.007	(0.005)
Emotion - fear							0.0004	(0.002)
Emotion - happy							0.004***	(0.001)
Emotion - sad							0.0003	(0.002)
Emotion - surprise							-0.002	(0.003)
Race - asian							0.002	(0.003)
Race - black							-0.002	(0.004)
Race - indian							-0.00003	(0.006)
Race - latino_hispanic							-0.0002	(0.005)
Race - middle_eastern							0.002	(0.004)
Facial attractiveness							0.001	(0.005)
Makeup heaviness							0.003	(0.017)
Object – top/tshirt/sweat							-0.002	(0.003)
Object – jacket							-0.085***	(0.025)
Object – pants							0.003	(0.006)
Object – skirt							0.050	(0.060)
Object – dress							0.003**	(0.001)
Object - glasses							0.0001	(0.006)
Object – hat							0.037	(0.038)
Object – watch							0.015	(0.030)
Object – belt							0.024	(0.036)
Object – tights/stockings							0.018	(0.016)
Object – shoe							0.002	(0.001)
Object – bag/wallet							-0.008	(0.020)

Object – collar							-0.028	(0.021)
Object – lapel							0.024*	(0.014)
Object – sleeve							-0.001	(0.001)
Object – pocket							0.031**	(0.014)
Object – neckline							0.001	(0.001)
Foundation1 (G)							0.002	(0.005)
Foundation2 (G)							0.003	(0.006)
Foundation3 (G)							0.003	(0.005)
Blush1 (G)							-0.002	(0.009)
Blush2 (G)							-0.004	(0.014)
Blush3 (G)							-0.017	(0.019)
Blush4 (G)							0.009	(0.023)
Blush5 (G)							0.015	(0.015)
Lip (G)							0.001	(0.004)
Lipliner (G)							-0.007	(0.007)
Eyeshadow1 (G)							-0.003	(0.005)
Eyeshadow2 (G)							0.001	(0.010)
Eyeshadow3 (G)							-0.004	(0.014)
Eyeshadow4 (G)							-0.020	(0.017)
Eyeshadow5 (G)							0.019	(0.012)
Eyeliner (G)							-0.011**	(0.005)
Eyebrow (G)							0.008*	(0.004)
NIMA score average							-0.002	(0.002)
Constant	0.570***	(0.017)	0.674***	(0.035)	0.674***	(0.035)	0.677***	(0.039)
Observation	101,512 (1,017)		101,512 (1,017)		101,512 (1,017)		101,512 (1,017)	
Fixed effects	Individual & Time		Ind/Time/Br/Type		Ind/Time/Br/Type		Ind/Time/Br/Type	
R^2	0.2942		0.3249		0.3262		0.3266	
Notes. *p < 0.1; **p < 0.05; ***p < 0.01.								

OA 6.I. ROBUSTNESS CHECK: AD SPENDING BY MEDIA SOURCE

In the IV model, we accounted for unknown shocks from brands' advertisements by incorporating ad spending at the brand-month level. However, brands may run several campaigns through different media channels, creating media-specific shocks, so we re-estimate the models with separate ad-spending variables for six media channels. In Table S45, the estimated coefficients are consistent with the main models: significantly positive for average loudness, and significantly negative for sponsorship and for sponsorship x average loudness.

Table S45. Results of Sponsorship & Voice Effects with Media-Specific Ad Spending

	Dependent Variable: Consumer Sentiment			
	IV		**FRD**	
Vocal characteristics	**Coef.**	**SE**	**Coef.**	**SE**

Average loudness	**0.0027****	**(0.0013)**	**0.0117****	**(0.0050)**
Average pitch	**0.0061**	**(0.0039)**	**0.0156**	**(0.0158)**
Loudness variability	**-0.0023**	**(0.0017)**	**-0.0108****	**(0.0055)**
Pitch variability	**-0.0025**	**(0.0032)**	**-0.0141**	**(0.0110)**
Talking duration	**-0.0088****	**(0.0039)**	**-0.0330****	**(0.0139)**
Sponsorship & Sponsored Vocal Characteristics				
Sponsorship	**-0.0058****	**(0.0027)**	**0.0041**	**(0.0148)**
Sponsorship * Average loudness	**-0.0053****	**(0.0025)**	**-0.0187****	**(0.0087)**
Sponsorship * Average pitch	**-0.0072**	**(0.0068)**	**-0.0230**	**(0.0182)**
Sponsorship * Loudness variability	**-0.0028**	**(0.0031)**	**0.0146**	**(0.0090)**
Sponsorship * Pitch variability	**0.0010**	**(0.0056)**	**0.0275***	**(0.0151)**
Sponsorship * Talking duration	**0.0046**	**(0.0069)**	**0.0353***	**(0.0180)**
Advertisement Expenses by Media				
Monthly ad. spending by Internet	-0.0010	(0.0007)	-0.0039	(0.0045)
Monthly ad. spending by magazines	0.000003	(0.0004)	-0.0001	(0.0022)
Monthly ad. spending by newspaper	0.0009	(0.0010)	0.0037	(0.0056)
Monthly ad. spending by others	-0.0001	(0.0015)	-0.0073	(0.0079)
Monthly ad. spending by radio	0.00005	(0.0004)	0.0027	(0.0023)
Monthly ad. spending by TV	0.0002	(0.0003)	0.0013	(0.0020)
Constant	0.7590***	(0.0845)	1.1489***	(0.3031)
Notes *p < 0.1; **p < 0.05; ***p < 0.01.				

REFERENCES

Brysbaert, Marc, Amy B. Warriner, and Victor Kuperman (2014), "Concreteness ratings for 40,000 generally known English word lemmas," *Behavior research methods*, 46 (3), 904-911.

Caliendo, Marco, and Sabine Kopeinig (2008), "Some practical guidance for the implementation of propensity score matching," *Journal of economic surveys*, 22 (1), 31-72.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi (2012), "Plausibly exogenous," *Review of Economics and Statistics*, 94 (1), 260-272.

Dave, Namrata (2013), "Feature extraction methods LPC, PLP and MFCC in speech recognition," *International journal for advance research in engineering and technology,* 1 (6), 1-4.

DiPrete, Thomas A., and Markus Gangl, (2004) "Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments," *Sociological methodology*, 34 (1), 271-310.

Eyben, Florian, Martin Wöllmer, and Björn Schuller (2010), "Opensmile: the munich versatile and fast open-source audio feature extractor." *Proceedings of the 18th ACM international conference on Multimedia.*

Eyben, Florian (2015), Real-time speech and music classification by large audio feature space extraction, *Springer*.

Green, Welley (2003), Economic Analysis, *Pearson Education India*.

Jacob, Robin, et al. (2012), “A practical guide to regression discontinuity.” *MDRC.*
Kießling (1997), Extraktion und Klassifikation prosodischer Merkmale in der automatischen Sprachverarbeitung, *Shaker* (ISBN 978-3-8265-2245-1)
Kyle, Kristopher, and Scott A. Crossley (2015), “Automatically assessing lexical sophistication: Indices, tools, findings, and application,” *Tesol Quarterly,* 49 (4), 757-786.
Kyle, Kristopher, Scott Crossley, and Cynthia Berger (2018), “The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0,” *Behavior research methods*, 50 (3), 1030-1046.
Lee, David S., and David Card (2008), “Regression discontinuity inference with specification error, ” *Journal of Econometrics*, 142 (2), 655-674.
Lee, David S., and Thomas Lemieux (2010), “Regression discontinuity designs in economics,” *Journal of economic literature*, 48 (2), 281-355.
McLoughlin, Ian Vince (2008), “Line spectral pairs,” *Signal processing*, 88 (3), 448-467.
Mediakix (2019) (Accessed October 7, 2019), https://mediakix.com/influencer-marketing-resources/influencer-tiers/
Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). An overview of the global historical climatology network-daily database. Journal of atmospheric and oceanic technology, 29(7), 897-910.
Mimno, David, et al. (2011), “Optimizing semantic coherence in topic models,” *Proceedings of the 2011 conference on empirical methods in natural language processing*.
Mittal, Anish, Anush Krishna Moorthy, and Alan Conrad Bovik (2012), “No-reference image quality assessment in the spatial domain," *IEEE Transactions on image processing 21 (12)*.
Montaguti, Elisa, Scott A. Neslin, and Sara Valentini (2016), “Can marketing campaigns induce multichannel buying and more profitable customers? A field experiment,” *Marketing Science*, 35 (2), 201-217.
Moorthy, Anush Krishna, and Alan Conrad Bovik (2011), “Blind image quality assessment: From natural scene statistics to perceptual quality,” *IEEE transactions on Image Processing,* 20 (12), 3350-3364.
Narwaria, Manish, and Weisi Lin (2010), “Objective image quality assessment based on support vector regression,” *IEEE Transactions on Neural Networks*, 21 (3), 515-519.
Newman, David, et al. (2010), “Automatic evaluation of topic coherence,” *Human language technologies: The 2010 annual conference of the association for computational linguistics.*
Papies, Dominik, Peter Ebbes, and Harald J. Van Heerde (2017), “Addressing endogeneity in marketing models,” *Advanced methods for modeling markets.* Springer, Cham, 581-627.
Röder, Michael, Andreas Both, and Alexander Hinneburg (2015), “Exploring the space of topic coherence measures,” *Proceedings of the eighth ACM international conference on Web search and data mining*.
Rosenbaum, Paul R (2002), “Overt bias in observational studies." *Observational studies.* Springer, New York, NY, 71-104.
Rosenbaum, Paul R., and Donald B. Rubin (1983), “The central role of the propensity score in observational studies for causal effects,” *Biometrika*, 70 (1), 41-55.
Rossi, Peter E (2014), “Even the rich can make themselves poor: A critical examination of IV methods in marketing applications,” *Marketing Science*, 33 (5), 655-672.

Schneider, Matthew J., and Sachin Gupta (2016) "Forecasting sales of new and existing products using consumer reviews," *International Journal of Forecasting*, 32 (2), 243-256.
Simpson A. P. (2009), "Phonetic differences between male and female speech," *Lang. and Linguistics Compass*, 3 (2), 621-640.
Talebi, H., & Milanfar, P. (2018). "NIMA: Neural image assessment." *IEEE transactions on image processing*, 27(8), 3998-4011.

2. Metaverse Is Near: The Impact of Virtual Influencers

ABSTRACT

Brands are increasingly choosing virtual influencers for marketing because they are cheaper than human influencers and exempt from human misconduct (e.g., moral issues, feuds, scandals). Despite the rapid growth in virtual influencers, however, little is known about whether virtual influencers complement or displace human influencers. In this paper, we use Instagram post data from 2011 to 2020 to study how the entry of virtual influencers affected the sponsorship deals received by human influencers and how human influencers strategically responded to the existential threat. We use deep learning techniques to measure potential confounding factors including influencer demographics (DeepFace for predicting age, gender and race), facial attractiveness (ResNet-55 architecture), and content topics (LDA-BERT Autoencoder). Then, we employ three identification strategies: doubly robust difference-in-differences, difference-in-differences, and inverse probability weighting. We find that, among the brands that employed virtual influencers, human influencers who previously were sponsored by those brands tended to be displaced, while human influencers who previously were *not* sponsored by those brands tended to be complemented by the introduction of virtual influencers. Vulnerability to displacement was greatest for older influencers, male influencers, and less-attractive influencers; brands in experience goods categories did not embrace virtual influencers as much as brands in other categories. We find evidence of engagement-based mechanisms for the average and heterogeneous treatment effects. Finally, human influencers responded to the introduction of virtual influencers by increasing their usage of human oriented verbs. We provide practical implications for the government, firms, and influencers.

1. Introduction

In 2018, Time magazine announced the 25 most influential people on the internet.[11] Surprisingly, included among obvious choices like Rihanna and former President Trump, was a Computer-Generated Imagery (CGI) "virtual influencer": Lil Miquela. Miquela Sousa, well-known for her Instagram nickname Lil Miquela, is a fictional character with over 3.1 million Instagram followers. Miquela is one of the successful creations of Brud, an AI and robotics startup company that creates digital characters and shares their stories on social media. Brud is now worth $125 million and has received more than $5 million in funding from investors such as Sequoia Capital and BoxGroup since the company's 2016 debut.[12] Miquela has been featured in product endorsements for fashion brands like Calvin Klein and Prada, and Miquela's contemporaries[13] are collaborating with a wider variety of brands. For instance, the virtual KFC influencer Colonel Sanders had a paid partnership with Turbotax[14] (example in Online Appendix 1), and virtual influencer Imma posted sponsored ads for Porsche.

Virtual influencers represent a risk-management strategy for brands as human influencers intermittently are blamed for deceptive marketing behaviors. For instance, Amazon Inc. sued 13 human influencers who promoted counterfeit luxury fashion goods,[15] and PR Consulting Inc. filed a lawsuit against Luka Sabbat for failing to complete most components of his $60,000 promotion contract with the client Snap Inc.[16] Beyond eliminating the risk of human misconduct,

[11] https://time.com/5324130/most-influential-internet/
[12] https://shanebarker.com/blog/cgi-influencers/
[13] Virtual Humans (https://virtualhumans.org/), a well-known virtual influencer tracking company, listed 144 virtual influencers at the time of writing.
[14] This project is in collaboration with the Facebook CrowdTangle team to extract Instagram posts and metadata.
[15] https://www.theverge.com/2020/11/12/21562758/instagram-influencer-counterfeit-gucci-dior-amazon-fake-listing
[16] https://www.influencerintelligence.com/blog/vh/what-luka-sabbats-breach-of-contract-means-for-the-future-of-influencer-marketing).

virtual influencers may appeal to brands because they are more reasonably priced,[17] do not experience physical, economic, or psychological breakdowns due to external shocks (e.g., the COVID-19 pandemic; Ong 2020), maintain their best appearance (since they do not age), and enable brands to deliver highly tailored messages.

Although these qualities make virtual influencers increasingly attractive to brands, no research has yet examined the impact of virtual influencers on the economic well-being of human influencers. We ask three research questions. First, do virtual influencers displace or complement human influencers? Put differently, do brands hire more or fewer humans as they begin to adopt virtual influencers? AI automation can increase productivity (positive complement effect), but it also can also replace human tasks (negative displacement effect; Acemoglu and Restrepo, 2018). We answer this question in a non-traditional social media content industry where the finished good differs across influencers. Second, which human influencers are hurt most and least by the adoption of virtual influencers? We consider heterogeneity based on the influencer's demographics, appearance, and tenure (years of experience) as well as the brand's product category (experience goods or not). Lastly, how do human influencers change their posting behaviors after the arrival of virtual influencers? The last question is derived from a unique feature of the content industry we study: content producers can change their creation strategies at will instead of making identical goods in the mass-production industry.

We answer our research questions by identifying how the introduction of virtual influencers in 2016 affected the number of sponsorships received by human influencers and the

[17] Lil Miquela has over 3 million followers and costs $8,500 per sponsored post, while human influencers with over 1 million followers commonly earn at least $10,000 per post (Ong 2020). Miquela is recorded as the highest earner of the virtual influencers.

number of human influencers sponsored by brands. We define *treated brands* as those that sponsored virtual influencers after their introduction to Instagram, while *control brands* are those that did not sponsor virtual influencers during the sample period (which ended in 2020). Then, *treated influencers* are those who were sponsored by treated brands before 2016, while *control influencers* were not. We use deep learning methods to detect potential confounds such as attractiveness and content topics, and we construct matched sets of influencers and brands. We use a two-level identification strategy by comparing sponsorships received by treated vs. control influencers in the post-treatment period as well as sponsorships given by treated vs. control brands in the post-treatment period. We construct a doubly robust difference-in-differences (DR-DiD) estimator to avoid the model misspecification bias (Athey and Wager, 2017; Imbens, 2020; Callaway and Sant'Anna, 2021), and we confirm the findings with a difference-in-differences (DiD) estimator and inverse propensity weighting (IPW).

We find that *treated* human influencers experienced a 16.81% decrease in the number of sponsoring brands after the introduction of virtual influencers. Surprisingly, *control* human influencers experienced a 10.63% increase in sponsoring brands, and treated brands increased their hiring of human influencers overall by 18.53% (through employing more control human influencers). The results validate both the AI displacement effect and complementarity effect: virtual influencers displaced treated human influencers while complementing control human influencers (Acemoglu and Restrepo, 2018; Brynjolfsson et al., 2018; Grace et al., 2018; Webb, 2020).

Next, we identify several characteristics that seem to make human influencers especially vulnerable to displacement from virtual influencers. Consistent with the automation potentials predicted by Arntz et al. (2017), older treated influencers lost 9.43% more sponsorships than their younger counterparts, and male treated influencers lost 6.48% more sponsorships than their female counterparts. We find an attractiveness premium, consistent with the literature (Hosoda et al., 2003): more-attractive treated influencers suffered a decrease in sponsorships of only 18.78%, while their less-attractive counterparts suffered a decrease of 56.79%. In a departure from the literature on heterogeneity in traditional job applicants (Kaufmann et al., 2016), we find that influencers who appeared older than their chronological age fared better than influencers who appeared younger than their age—older-looking treated influencers suffered a decrease in sponsorships of only 2.08%, while their younger-appearing counterparts suffered a decrease of 10.86%. Influencer tenure did not moderate the treatment effect, consistent with Colombo et al. (2019).

Similarly, not all brands were equally likely to embrace virtual influencers as viable substitutes for human influencers. We hypothesize that brands in experience goods categories might be more reluctant to use virtual influencers because consumers are known to eschew AI in domains that seem especially human (Luo et al., 2019; Longoni et al., 2019). As we predicted, treated brands in the experience goods category increased the number of sponsored human influencers more than treated brands in other product categories in the post-treatment period.

Finally, we find that treated human influencers responded to the threat of virtual influencers by using more human-specialized (human-oriented) verbs, categorized by VerbNet (Schuler 2005) as verbs that represent *help, future having, consume, grow, wink, stalk, and complain*. The significant shift in posting behavior provides evidence of a new Luddite movement to avoid job displacement and showcase the importance of human workers in the influencer market.

2. Related Literature

Our theory is derived from and contributes to research on (i) the complementarity and displacement effects of AI on human labor, (ii) pushback from the demand-side (consumers or firms that use AI technologies), and (iii) heterogeneity in job market successes by job and employee characteristics. Methodologically, our paper follows the literature on the robust estimator methods (e.g., DR-DiD) that are appropriate for circumstances with a risk of model misspecification.

2.1. AI Complementarity vs. Displacement Effects on Human Labor

Previous literature has speculated about how AI may both complement and displace human employment. Frey and Osborne (2017) predict that job categories involving originality, persuasion, and social perceptiveness are less susceptible to replacement. Arntz et al. (2017) optimistically argue that the automation risk drops to 9% after accounting for non-automatable tasks such as consulting and training. Similarly, Felten et al. (2018) and Colombo et al. (2019) suggest that automation will lead to few job displacements and will complement more jobs than it will replace. By contrast, Grace et al. (2018) predict that AI will replace human occupations such as translating languages by 2024, driving trucks by 2027, and retail by 2031. Brynjolfsson et al. (2018) construct the suitability for machine learning (SML) measure and find that most occupations are at some risk of replacement by automation. Acemoglu and Restrepo (2018) and Webb (2020) develop a conceptual framework of how AI replaces human labor. On the one hand, AI automation substitutes cheaper capital for expensive human labor, so it raises the firm's productivity and demand for human labor in non-automatable tasks (positive complementarity effect; Acemoglu and Restrepo, 2018). On the other hand, automation reduces the number of jobs available for humans and may suppress human wages as the human tasks are subject to diminishing returns (negative displacement effect; Acemoglu and Restrepo, 2018).

The existing findings focus on industries that do not involve creativity—the finished products are identical regardless of whether the manufacturer is a human or AI (e.g., conveyor belt manufacturing system), so humans and machines are perfect substitutes at the task level. Moreover, the datasets are limited to job recruitment details, such as the O*NET job description database. Table 1 compares our paper with the existing literature.

Table 1. Comparison of the Literature on the Impacts of AI on the Human Labor Market

Author(s)	(1) Can producers change their final products after AI disruption?	(2) Heterogeneity	(3) Datasets
Arntz et al. (2017)	No	Demographics, skills, job category	Survey of Adult Skills (PIACC)
Frey and Osborne (2017)	No	Job characteristics	O*NET job description database
Grace et al. (2017)	No	Beliefs about progress in AI	Survey of AI researchers
Brynjolfsson et al. (2018)	No	Job characteristics	O*NET job description database
Acemoglu and Restrepo (2018)	No	-	-
Felten et al. (2018)	No	Job characteristics	O*NET job description database
Colombo et al. (2019)	No	Job characteristics	Wollybi[18] job description database
Webb (2020)	No	Patents and job characteristics	O*NET database
Our Paper	Yes (Influencers can change their content after the disruption)	Demographics, facial attractiveness, user engagements	Instagram posts, Influencer profile databases

Notes. The first question, whether workers (producers) can change their final products, is comparable to the question of whether humans and AI produce the same finished products (i.e., no creativity is needed). In the influencer content business, every influencer's content is distinct from their peers'. The creativity involved in the final product separates our paper from all others in the table. Acemoglu and Restrepo (2018) is a theory paper, so it does not entail datasets.

[18] Wollybi is an Italian job posting and description website.

2.2. Demand-Side Pushback Effects

Our paper also expands the understanding of how consumers and firms push back against AI adoption. When algorithms were introduced for forecasting in the early 2000s, many firms refused to use them even though reliance on quantitative estimates from algorithms led to fewer forecasting errors (Sanders and Manrodt, 2003). Even many years later, professionals were reluctant to give sufficient weight to forecasts from algorithms (Fildes and Goodwin, 2017). Similarly, consumers become less willing to purchase once they recognize that they are interacting with an AI chatbot instead of a human (Luo et al., 2019), and consumers have lower reservation prices for healthcare provided by AI (Longoni et al., 2019). Luo et al. (2019) infer that such negative responses might reflect the subjective human perception that AI lacks empathy. Likewise, Longoni et al. (2019) find that resistance to medical AI is stronger for people who perceive themselves as more unique, such that algorithms might not be able to understand them. Based on consistent findings of resistance to AI among firms and consumers alike, we hypothesize that human influencers who are threatened by virtual influencers will strategically change their posting behaviors to exhibit their humanness. We test for this possibility by investigating changes in verb usage, as some verbs are more human-specific or human-oriented than others.

2.3. Job Market Success by Demographics, Appearance, Tenure, and Product Category

We investigate how the impact of virtual influencer debuts may vary with influencer and brand heterogeneity. The literature documents variation in job market success on the basis of demographics, appearance, tenure, and job category. Regarding the risk of replacement by AI, Arntz et al. (2017) predict that older, male, and less-educated workers are at higher risk of being replaced by automation in the US;[19] Colombo et al. (2019) find no difference by the years of worker experience (i.e., tenure). Sparse literature has examined how the worker's appearance influences whether AI will be complementary or substitutive, though appearance is known to affect other domains of job-related success. Hosoda et al. (2003) show that a physical attractiveness bias influences job-related outcomes even when the applicant's job-relevant information is known. Also, job candidates who appear older (vs. the same as or younger) than their chronological age are perceived as less healthy and are less likely to be hired (Kaufmann et al., 2016). Finally, we compare treatment effects between brands that are vs. are not in experience goods categories, which we expect should be less amenable to virtual influencers because of the inherent humanness of the products.

2.4. Doubly Robust Estimation of the Average Treatment Effect (ATE)

We employ the doubly robust (DR) estimator because it is robust to model misspecification (Imbens, 2020). Brands are known to select influencers for sponsorship based on popularity (consumer engagement with posts) and demographics (Hwang et al., 2021), but it is not known how exactly they do so (e.g., how some brands choose popular female influencers while other brands prefer their male counterparts). The DR estimator involves estimating both the conditional expectations of potential outcomes and the propensity score, rendering it less sensitive to estimation error in either (Athey and Wager, 2017). As our question is answered in a panel dataset, we employ the DR-DiD estimator (Callaway and Sant'Anna, 2021). To the best of our knowledge, our paper is among the first to implement a "doubly robust" estimation methodology in business applications.

[19] In Arntz et al. (2017), Table 1 exhibits several determinants of the risk of replacement by automation in the US: worker characteristics (gender, age, education), skills (literacy, numeracy, problem-solving), job characteristics (e.g., firm-size, income), and tasks (e.g., selling, consulting).

3. Data

We use Instagram post data created by virtual and human influencers.[20]

3.1. Create a List of Virtual Influencers

We start with a list of virtual influencers and then identify relevant brands and human influencers. To create the list of virtual influencers, we collaborate with VirtualHumans,[21] a company that publishes news, research, interviews, and biographies about virtual influencers. As of April 2020, VirtualHumans had information on 136 virtual influencers: 19 (14%) were male human-like, 87 (64%) were female human-like, and 30 (22%) were not human-like (e.g., animals, aliens, babies, and 2D characters), as shown in Figure 1. Also, 20 of the listed virtual influencers were owned by a brand, while the remaining 116 could work freely with multiple brands.

Figure 1. Male, Female, and Other Types of Virtual Influencers

<table>
<tr><td colspan="2">Male Virtual Influencer
(Example: Phoenix McEwan)</td><td colspan="2">Female Virtual Influencer
(Example: Alice Mikoni)</td></tr>
<tr><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td colspan="4">Other Virtual Influencers</td></tr>
<tr><td>3D-Animated
Gender-Neutral
(Example: 8ubbles)</td><td>3D-Animated
Male
(Example: John Pork)</td><td>3D-Animated
Female
(Example: Ilona)</td><td>3D-Animated
Gender-Neutral
(Example: Kayda)</td></tr>
</table>

[20] Our Instagram post data and metadata were provided by the Facebook CrowdTangle team.
[21] VirtualHumans: www.virtualhumans.org

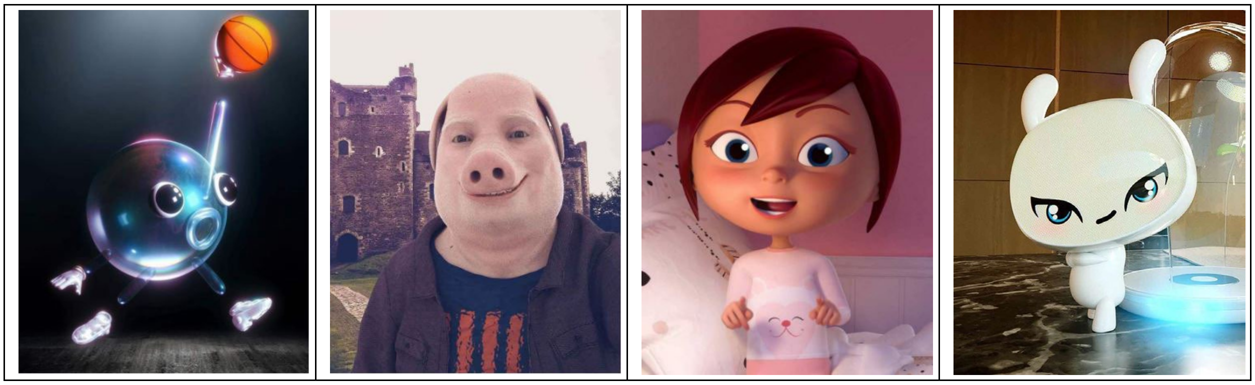


Table 2 provides the summary statistics of popularity metrics and posting behaviors of all 136 virtual influencers. We focus on the 52 virtual influencers (listed in Online Appendix 2, Table S1) that were sponsored by at least one brand during the sample period.

Table 2. Summary Statistics of Virtual Influencer Popularity and Posting Behaviors

Popularity Metrics	N	Mean	SD	Min.	Max.
Follower count (by influencer)	95	276,893.30	837,383.8	50	5,692,100
View count (by post)	31,603	16,762.39	433,211.9	0	53,193,884
Like count (by post)	31,603	13,226.22	336,723	0	54,904,361
Comment count (by post)	31,603	265.38	18,920.89	0	3,346,600
Posting Behaviors					
Launch year (by influencer)	136	2018.41	1.84	2011	2020
N. of posts (per year)	343	92.14	132.53	1	1,058
N. of sponsored posts (per year)	343	3.23	20.42	0	293
N. of sponsored posts (per brand)	287	15.58	213.19	1	3,614
Post text length (by characters)	31,603	185.47	196.06	0	2,191

Notes. The follower count includes missing values as our data does not keep track of the follower count of Instagram accounts with fewer than 50 followers. We calculate the number of posts per year as the influencer's total posts divided by the number of months since joining Instagram.

Among the 136 virtual influencers, 5 (4%) have more than 1 million followers on Instagram, 25 (18%) have 100,000 ~ 1 million followers, 34 (24%) have 10,000~100,000 followers, 29 (21%) have 1,000~10,000 followers, and 44 have fewer than 1,000 followers. The virtual influencers were sponsored by a total of 287 brands (listed in Online Appendix 2, Table S2). The brands sponsored 15.58 posts on average in their partnerships with virtual influencers. Top sponsoring brands include Yoox (3,614 posts), luxury brands such as Dior (30 posts), Versace (28 posts), and Valentino (24 posts), and experience good brands (e.g., Airbnb, Netflix, Tinder, Toysrus, and Uber).

3.2. Sponsorship Disclosure

The Federal Trade Commission (FTC) mandates that online influencers must disclose sponsorships with simple and clear language, so we relied on sponsorship disclosure to identify sponsored relationships between influencers and brands. Figure 2 displays two examples, and Table 3 lists 10 types of language that influencers might use.[22]

Figure 2. Examples of Sponsorship Disclosure Language

Language: #ad	Language: Partnered up with "Brand"
#ad Enjoying a sip of my favourite @drinkmashup flavour with white tea and passion fruit before my combo.	Dry January continues! I partnered up with @drysoda to make this delicious alcohol free cocktail you'll love!

Table 3. Common Types of Sponsorship Disclosure Language

No.	Sponsorship Disclosure Language	Similar Languages
1.	#ad	#ad, #adv, #advertisement
2.	Partnered with "Brand"	Partnered up (partnership) with "Brand"
3.	Sponsored by "Brand"	"Brand" is sponsoring today's post
4.	Thank you "Brand" for sponsoring	Thanks "Brand" for sponsoring
5.	Collaborated with "Brand"	Collaboration (collab) with "Brand"
6.	Teamed up with "Brand"	Teaming up with "Brand"
7.	Supported by "Brand"	"Brand" supports this post
8.	Powered by "Brand"	This post is empowered by "Brand"
9.	A "Brand" ambassador	I become a virtual ambassador of "Brand."
10.	An advertisement for "Brand"	This content is an advertisement of "Brand"
Notes. The first column lists the most representative form of the sponsorship disclosure language with distinctive root words. The third column lists derivatives in the same category		

[22] For detailed information, please visit the FTC's influencer guidebook: https://www.ftc.gov/system/files/documents/plain-language/1001a-influencer-guide-508_1.pdf

of disclosure language. For the proportions of each sponsorship tag, please see Online Appendix 2, Table S3.

Instagram also implemented its own enforcement tool (rule) on March 30, 2017, to improve the transparency and consistency of branded content campaigns on Instagram.[23] The tool enables influencers to tag their business partners (and Instagram requires that they do so), as shown in Figure 3.[24] We use the tags as well as sponsorship disclosure language to identify sponsorships.

Figure 3. Examples of Branded Content Disclosure with Instagram Tag Tool

Example of Human Influencer Post with Tag	Example of Virtual Influencer Post with Tag
	

3.3. Create the Treatment and Control Groups

Our *treated brands* are those that sponsored virtual influencers after 2016, when Lil Miquela debuted, and also sponsored human influencers before 2016. As shown in Figure 4, our *treated influencers* are those who were sponsored by treated brands before 2016, such that they faced

[23] Branded content tools on Instagram. https://business.instagram.com/a/brandedcontentexpansion
[24] Each tagged business partner can keep track of the reach and engagement (likes, comments) of the tagged posts. For detailed information, please visit https://business.instagram.com/a/brandedcontentexpansion.

direct consequences—be it a positive complementarity effect or negative displacement effect—from the introduction of virtual influencers. For more information on data collection, see Online Appendix Section 2.

Figure 4. Process of Identifying Treated Human Influencers & Treated Brands

Collect the Instagram posts of all AI virtual influencers

↓

From the posts, identify the "AI sponsors," that is, all sponsors of the virtual influencers

↓

List the human influencers sponsored by the AI sponsors

↓

Define the "treated influencers" as the humans sponsored by AI sponsors before 2016

↓

Define the "treated brands" as the AI sponsors that hired both virtual and human influencers before 2016

To determine whether the introduction of virtual influencers had a predominantly complementary effect (Arntz et al., 2017; Felten et al., 2018; Colombo et al., 2019) or displacement effect (Acemoglu and Restrepo, 2018; Brynjolfsson et al., 2018; Grace et al., 2018; Webb, 2020), we need to compare the treated influencers with control influencers: those who were *not* sponsored by the treated brands before 2016, but who resemble the treated influencers in other regards. We identify eligible control influencers as those who used a treated brand's products or included a treated brand's name in their Instagram content (but did not have any sponsored posts with a treated brand before 2016). To ensure that treated and control influencers are indeed *influencers*, we restrict the sample to those listed on famousbirthdays.com, a website that catalogs the biographies of social media influencers and TV celebrities. We identify each influencer's age and gender from the website.

Then, we use PSM (equation in Section 4.1.1) to construct the final sample of control influencers. Influencer and post characteristics may affect brands' sponsorship decisions, so we incorporate both types of attributes to create a set of control influencers who closely resemble the treated influencers in their likelihood of being treated. Specifically, we include demographic characteristics (age, gender, and race), appearance (facial attractiveness, predicted age), tenure, post characteristics (11 topics and the text length), user engagement (likes and comments), posting behaviors (the number of sponsored and total posts), and time fixed effects (year and month). In the next section, we explain how we collected these variables.

3.4. Control Variables

3.4.1. Demographic Characteristics

We acquired each human influencer's gender and chronological age from famousbirthdays.com, as mentioned in the previous section. Then, we use the DeepFace (Taigman et al., 2014), a deep-learning-based facial attribute analysis framework from the Facebook research team, to predict each human influencer's race (six categories: Asian, Black, Indian, Latino/Hispanic, Middle Eastern, or Non-Hispanic White), to measure their facial attractiveness, and to predict their age based on appearance. Please see Online Appendix 3.A for details.

For each virtual influencer, we collected the gender from VirtualHumans,[25] and we used the DeepFace model to predict their age and race. Figure 5 displays two examples.

Figure 5. Predicted Age, Gender, and Race of Virtual Influencers[26]

Name: Liam Nikuro (@liam_nikuro)	Name: Ella (@ella.imagination)

[25] VirtualHumans.org provides lists of masculine and feminine virtual influencers, so we use the list to classify each virtual influencer as male or female.

[26] Liam_nikuro is known to be Japanese-American (https://www.scmp.com/sport/basketball/article/3098800/what-liam-nikuro-and-how-did-he-get-nba-bubble), and Ella.imagination is known to be Asian (https://aww.tokyo/news/2020/10/254/).

- Predicted Age: 25.60
- Gender: Male
- Race: White (35%)
  Latino/Hispanic (23%)
  Asian (18%)
  Indian (12%)
  Middle Eastern (8%)
  Black (4%)

- Predicted Age: 29.37
- Gender: Female
- Race: Asian (76%)
  White (22%)
  Latino/Hispanic (1%)
  Indian (0%)
  Middle Eastern (0%)
  Black (0%)

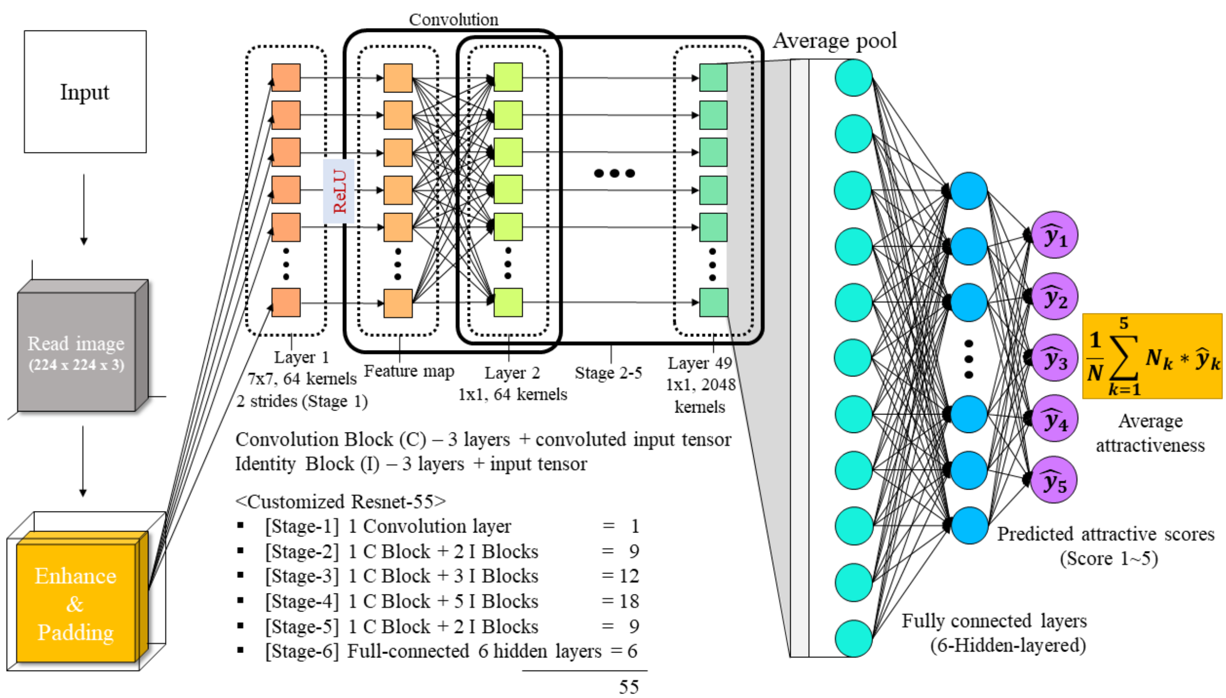
3.4.2. Influencer Attractiveness Using a Customized Resnet-55 Model

We match influencers on facial attractiveness because attractiveness is known to help influencers get more sponsorship deals in the influencer labor market (Baker and Churchill, 1977). We use the SCUT-FBP5500 dataset (Liang et al., 2018) to train a ResNet-55 model. Before applying the trained model to our influencer dataset, we debias the photos (which were taken under varying conditions) for rotation, brightness, contrast, and enhancement, following the image hashing and enhancement literature (Tang et al., 2013). Then, we train our Resnet-55 model with ImageNet weights based on the original deep residual learning model (He et al., 2015).[27] Our Resnet-55 model outperforms the original residual network model proposed by Liang et al. (2018), with 0.1087 as the Mean Absolute Error (MAE) loss value and 0.1477 as the Rooted Mean Squared

[27] We follow the empirical experiment results from the Microsoft research team (He et al., 2015).

Error (RMSE) loss value.[28] For the convergence plot, see Online Appendix 3.B. Figure 6 illustrates the architecture of our attractiveness recognition algorithm via the Resnet-55 model.

Figure 6. Architecture of Image-Enhancement and Resnet-55 Model Algorithm



3.4.3. Topic Characterization Using Autoencoder Based on LDA+BERT Embeddings

To control for the content of the influencer's posts (as well as metrics such as text length), we combine Latent Dirichlet allocation (LDA; Blei et al., 2003) with Bidirectional Encoder Representations from Transformers (BERT) embeddings (Devlin et al., 2018). We do not use LDA alone because the short texts of social media posts do not have sufficient word occurrences (Sriram et al., 2010).[29] Instead, we implement an L1-regularized autoencoder with three convoluted layers to represent the latent word space based on the concatenated LDA and BERT embeddings. The LDA-BERT L1-Regularized AutoEncoder model outperforms the traditional

[28] In Liang et al. (2018), the benchmark performance measures (MAE and RMSE) are 0.2291 and 0.3017. Based on the paper's reported accuracies, our model outperforms compared to the reported accuracy scores.
[29] In our analysis, the average coherence score achieved by LDA was 20~30%.

LDA and BERT models on short social media text data, and it parsimoniously characterizes content through L1 norm regularization. Figure 7 shows the topic characterization process. Our final model contains 11 topics because this version achieved the highest coherence score (see Online Appendix 3.C for details). Table 4 provides the top 10 words for each topic.

Figure 7. Architecture of LDA-BERT L1-Regularized AutoEncoder (AE) Model

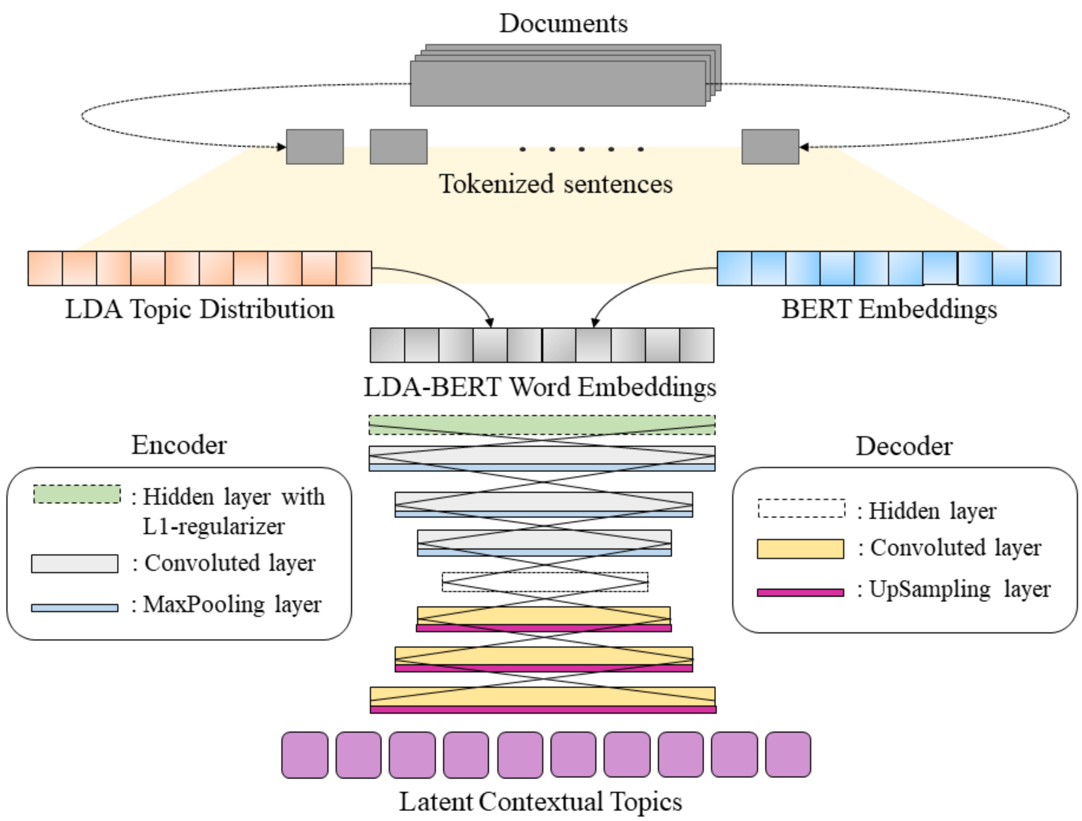


Table 4. 11 Topics and Top 10 Words

Topic	Word
Topic 0 - Routine 1	day, time, today, year, fit, thank, life, week, friend, la
Topic 1 - Fashion 1	shoe, fashion, dress, bag, today, boot, jacket, jean, style, look
Topic 2 - Social Event 1	thank, love, day, today, mem, time, night, fun, friend, show
Topic 3 - Cosmetics	chanel, cosmet, lip, eye, makeup, gucci, liner, lash, brow, label
Topic 4 - Social Event 2	night, day, show, week, today, time, parti, tonight, year, tomorrow
Topic 5 - Car	porsch, car, ferrari, vet, race, fraser, preseason, teakettl, today, day
Topic 6 - Others	mem, la, que, victoria, today, redbud, pow, basel, comfort, adida
Topic 7 - Fashion 2	fashion, thank, dress, hair, girl, beauti, makeup, love, style, vogu
Topic 8 - Routine 2	time, day, mem, today, life, peopl, thank, thing, redbud, way

Topic 9 - Fashion 3	fashion, vogu, com, show, pari, dress, style, hair, chanel, photo
Topic 10 - Streaming	netflix, amazon, itun, video, com, today, help, link, movi, book

3.4.4. User Engagement

Finally, we control for the influencer's popularity and engagement using the number of followers, comments, and likes. Table 5 provides summary statistics for the treated and control influencers.

Table 5. Summary Statistics of Influencer Popularity

Treated Influencer Popularity	N	Mean	Median	SD	Min.	Max.
Follower count (by year)	4,944	1,171,266	179,911.2	4,910,307	835.12	124,500,829
Like count (by year)	4,944	24,342.53	2,735.64	103,085.9	5.68	2,661,160
Comment count (by year)	4,944	203.91	46.36	644.49	0.19	16,196.31
Control Influencer Popularity	N	Mean	Median	SD	Min.	Max.
Follower count (by year)	4,797	769,906.2	152,348.2	2,124,498	1,193.76	49,586,360
Like count (by year)	4,797	18,463.07	2,800.49	80,505.24	0.6	2,215,818
Comment count (by year)	4,797	251.06	47.84	1,735.99	0.24	59,579.34
Notes. The like and comment counts are averaged at the influencer-year level in the annual panel data.						

4. Model

Our goal is to estimate the effect of introducing virtual influencers on the treated human influencers (i.e., those who previously were sponsored by a brand that hired one or more virtual influencers). We examine consequences on both the influencer and brand sides with two primary outcome variables: influencer i's number of sponsoring brands (Y_{it}), and brand k's number of sponsored influencers (Y_{kt}). The two-level analysis enables us to explain how brands' sponsorship decisions changed with the rise of virtual influencers (brand-level analysis) and how the changes in brand decisions affected human influencers (influencer-level).

The two-level model identification strategy also enables us to address the self-selection problem, as human influencers were not randomly assigned to treated and control brands. First, at the influencer level, we identify the treatment effect on humans who did (vs. did not) receive sponsorships from treated brands (i.e., those that hired virtual influencers after treatment and human influencers in the pre-treatment period). Second, at the brand level, we identify the treatment effect on brands that did (vs. did not) sponsor virtual influencers. In both analyses, we assume that we can control for the influencer's unobserved likelihood of being in the treatment group by conditioning on a rich set of observed characteristics. That is, we use a quasi-experimental matching procedure in which we match treated influencers (those who were sponsored by a treated brand before 2016) with similar control influencers (those who were sponsored by a treated brand only after 2016) based on a propensity score constructed from the influencer's characteristics and posting behaviors. In the brand-level analysis, we also match treated brands (those that sponsored virtual influencers) with control brands (those that did not sponsor virtual influencers) based on the observed brand characteristics used in Acemoglu et al. (2020).

In Section 4.1, we explain how we used PSM to construct the panel dataset of treated and control brands (4.1.1) and influencers (4.1.2). Section 4.2 explains how we used three identification strategies (DR-DiD, DiD, and IPW) to robustly confirm the ATE. Section 4.3 introduces the models for heterogeneous treatment effects (HTEs) by influencer demographics, appearance, and tenure as well as by the brand's product category (experience goods or not). Section 4.4 proposes a method for using consumers' engagement with posts to investigate the mechanisms underlying the ATE and HTEs. Lastly, Section 4.5 proposes a model for exploring changes in verb usage as an indication of strategic responses from human influencers who were threatened by the introduction of virtual influencers.

4.1. Treatment & Control Group Assignments with PSM

4.1.1. Brands

The *treated brands* are the 130 brands that sponsored virtual influencers between 2016 and 2020 and also sponsored human influencers before 2016. To construct a valid group of *control brands* (that had not sponsored virtual influencers as of 2020), we needed to identify brands that were similar to the treated brands on firm characteristics that may affect the decision to sponsor virtual influencers. We adapt variables from Acemoglu et al. (2020), who incorporated five firm-level outcomes as control variables to identify how firm performance was affected by firm-level robot adoption.[30] Table 6 compares the variables used in the two papers. Equation (2) is the matching equation.

We include all of the aforementioned characteristics because they yielded a better Akaike (Bayesian) information criterion value than alternative models with fewer sets of control variables (see Online Appendix 3.D). We matched the 130 treated brands with 130 control brands.

Table 6. Comparison of Control Variables in Acemoglu et al. (2020) and Our Paper

Classification	**Acemoglu et al. (2020)**	**Our Paper**
1. Employment	Log employment	Log of brand k's number of sponsored posts $\left(\overline{Sponsored\_posts_k}\right)$ in the pre-treatment period
2. Productivity	Log value added per worker	Log of the average user engagement (sum of likes and comments) that each influencer has acquired during the pre-treatment period ($\overline{Productivity_{k,t-1}}$)

[30] Acemoglu et al. (2020) used five dependent variables: (1) value added, (2) productivity, (3) labor share, (4) employment, and (5) wages.

3. Community (Brand Owned Media)	Fixed effects for the commuting zone that houses the firm's largest establishment	Log of the number of brand k's Instagram account followers ($Followers_k$), likes ($Likes_k$), comments ($Comments_k$), and posts per day ($Post\_per\_day_k$)
4. Large Corporation	Dummies for whether the firm is affiliated with a larger corporate group	Dummies for whether the firm is affiliated with a larger (parent) corporate group ($\overrightarrow{Large\_firm_k}$)
5. Industry	4-digit industry Standard Industrial Classification (SIC)	4-digit industry Standard Industrial Classification (SIC) code ($\overrightarrow{SIC\_code_k}$)

$$\begin{aligned} Treated\_brand_k &= \beta_0 + \beta_1 log(1 + Sponsored\_posts_k) \\ &+ \beta_2\, log(1 + \overline{Productivity_{k,t-1}}) + \beta_3\, log(1 + Followers_k) \\ &+ \beta_4\, log(1 + Likes_k) + \beta_5\, log(1 + Comments_k) + \beta_6 log\ (1 \\ &+ Post\_per\_day_k) + \vec{B}_7{}'\overrightarrow{Large\_firm_k} + \vec{B}_8{}'\overrightarrow{SIC\_code_k} + \epsilon_k \end{aligned} \tag{2}$$

4.1.2. Influencers

The *treated influencers* are the 669 human influencers who were sponsored by the 130 treated brands (described in Section 4.1.1) before 2016. In Equation (1), we use PSM to identify a group of *control influencers* who are similar to the treated influencers but did not have sponsorships from treated brands before 2016.

$$\begin{aligned} Treated\_influencer_i &= \alpha_0 + \alpha_1 Male_i + \alpha_2 \log\ (Age_i) + \vec{A}_3{}'\overrightarrow{Race_i} \\ &+ \alpha_4 \log\ (Attractiveness_i) + \alpha_5 \log\ (Predicted\_age_i) \\ &+ \alpha_6 Norm\_elapsed\_days_i + \vec{A}_7{}'\overrightarrow{Topics_i} + \alpha_8 \log\ (1 + Followers_i) \\ &+ \alpha_9 \log\ (1 + Likes_i) + \alpha_{10} \log\ (1 + Comments_i) + \alpha_{11} \log\ (1 \\ &+ Total\_posts_i) + \alpha_{12} \log\ (1 + Sponsored\_posts_i) + \epsilon_i \end{aligned} \tag{1}$$

where $Treatment_i$ is the binary treatment variable. We match on demographic characteristics ($Male_i$, a binary indicator of whether influencer *i* is male; Age_i, log-transformed; and $\overrightarrow{Race_i}$), appearance (log ($Attractiveness_i$), influencer *i*'s log-transformed attractiveness score from the ResNet model, and log ($Predicted\_age_i$), predicted by the DeepFace model), and tenure on Instagram ($Norm\_elapsed\_days_i$). We also match on post characteristics: the topics in influencer i's posts ($\overrightarrow{Topics_i}$), three log-transformed popularity metrics ($Followers_i$, $Likes_i$, and $Comments_i$), and two log-transformed numbers of posts ($Total\_posts_i$ and $Sponsored\_posts_i$).

As in the brand matching procedure, we chose the set of control variables that yielded the best Akaike (Bayesian) information criterion value (see Online Appendix 3.D). We matched 669 control influencers with the 669 treated influencers.

4.2. Identification Strategy

We follow the potential outcome (PO) framework (Rubin, 1974) to identify the ATE of the introduction of virtual influencers on human influencers and brands.[31] We estimate the ATE because the introduction of virtual influencers affected all market entities at the same time; we do not have a noncompliance issue.[32]

In Section 4.2.1, we use DR-DiD (Callaway and Sant'Anna, 2021) to estimate the treatment effect with a low risk of model misspecification (Athey and Wager, 2017; Imbens, 2020). Also, DR-DiD enables us to estimate time-variant treatment effects over multiple post-treatment periods, so we can gain insight into how the effect evolved. Then, in Section 4.2.2, we use DiD and IPW estimators as robustness checks. The DR-DiD method requires the same amount of reference periods for all influencers and brands, so we had to impute zeros for influencers and brands that did not have data as early as 2011 (the start of the reference period). The DiD and IPW estimators do not come with this limitation, so they help us confirm the robustness of the estimated treatment effects.

[31] Both the PO framework and causal Directed Acyclic Graphs (DAG) approach are widely used in the causal inference literature in marketing, economics and computer science. We chose the PO framework because it is appropriate for binary treatment effects (Imbens, 2020).
[32] The treatment variable is the introduction of virtual influencers, not whether the focal influencer was sponsored.

4.2.1. Doubly Robust Difference-in-Differences (DR-DiD) Estimator of the ATE

The DR-DiD estimator constructs an unbiased causal inference model by integrating the framework of potential outcome (PO) regression (Rubin, 1974), conditional on the treatment and the framework of the propensity score for the treatment effect. The DR-DiD estimator remains unbiased as long as either the regression or the propensity score model is correctly specified (Athey and Wager, 2017), so it is relatively robust to model misspecification and is used in the causal inference literature in economics, statistics, and mathematics (Imbens, 2020). We employ a version of the DR-DiD estimator that is suitable for more than two periods (Callaway and Sant'Anna, 2021)[33] so that we can characterize how the ATEs evolved during the post-treatment period. Following Callaway and Sant'Anna (2021), the DR-DiD specification to estimate the ATE is:

$$\begin{aligned} Y_{gt} = \beta_{0gt} + \beta_{1gt} \times 1\{G_g = 1\} + \beta_{2gt} \times 1\{\tau = t\} \\ + {\phi_{gt}}^{DR-DiD} \times 1\{G_g = 1\} \times 1\{\tau = t\} + \vec{\Gamma}' \times \vec{X} + \epsilon_{gt} \end{aligned} \tag{3}$$

$$\phi_{gt}^{DR-DiD} = E\left[Y_t - Y_{g-1} \middle| \vec{X}, G_g = 1\right]$$

$$= E\left[\left(\frac{G_g}{E(G_g)} - \frac{\frac{p_{gt}(\vec{X})C}{1 - p_{gt}(\vec{X})}}{E\left[\frac{p_{gt}(\vec{X})C}{1 - p_{gt}(\vec{X})}\right]}\right)\left(Y_t - Y_{g-1} - E\left[Y_t - Y_{g-1} \middle| \vec{X}, C = 1\right]\right)\right]$$

where Y_{gt} is the outcome of group g (the units whose first treatment period is time g) at time t. The introduction of virtual influencers occurred in 2016 for all influencers and brands, so g is the year 2016 for all influencers and brands.[34] ϕ_{gt}^{DR-DiD} is the ATE on the outcome in period t (Y_t) of the cohort that initially participated in treatment in period g ($G_g = 1$; G_g is a binary indicator of

[33] It also allows for variation in the timing of treatment (as staggered treatment adoption sometimes occurs in the econometrics literature).

[34] Callaway and Sant'Anna (2021) provide two-way estimation methods with different comparison groups such as the never-treated or not-yet-treated. Since our treatment period starts in 2016 for all entities, our comparison group is never-treated.

the start of treatment), relative to the cohort that never adopted ($C = 1$; C is 1 for the control group). In the ϕ_{gt}^{DR-DiD} formula, Y_t, Y_{g-1} are the treatment outcomes at time *t* and *g-1* (the last pre-treatment period), and $\vec{X}$ is the vector of influencer and content characteristics in the pre-treatment periods. $p_{gt}(\vec{X})$ is the propensity score, indicating the probability of being treated at time *g*, conditional on covariates $\vec{X}$. Formally, the treatment effects for influencer *i* and brand *k* are provided in Equation (4).[35] For more details about the DR-DiD estimator, see Online Appendix 4.A.

$$\phi_{gt,influencer}^{DR-DiD} = E\left[Y_{i,t} - Y_{i,g-1} \middle| \vec{X_i}, G_g = 1\right] \tag{4}$$
$$\vec{X_i} = [Male_i, \log(Age_{i,g-1}), \overrightarrow{Race_i}, \log(Attractive_i), \log(Predicted\_age_{i,g-1}),$$
$$Elapsed_{days_i}, \overrightarrow{Topics_i}, \log(1 + Followers_i), \log(1 + Likes_i),$$
$$\log(1 + Comments_i), \log(1 + Total\_posts_i)]$$

$$\phi_{gt,brand}^{DR-DiD} = E\left[Y_{k,t} - Y_{k,g-1} \middle| \vec{X_k}, G_g = 1\right]$$
$$\vec{X_k} = [log\left(1 + \overrightarrow{Productivity_k}\right), log(1 + Followers_k), \log(1 + Likes_i),$$
$$\log(1 + Comments), log\ (1 + Post\_per\_day_k), \overrightarrow{Large\_firm_k}, \overrightarrow{SIC\_code_k}]$$

4.2.2. Alternative ATE Model Specifications with DiD and IPW Estimators

The DR-DiD estimator is useful for accounting for time-variant treatment effects, but it requires all units to have the same amount of reference periods—and not all influencers or brands in our panel started working at the same time. In the DR-DiD estimation, we had to assume that influencers and brands existed even when they were not receiving (or giving) sponsorships on Instagram. The DiD and IPW estimators allow us to relax this assumption.

DiD Estimator

[35] $Engagements_i$ is the sum of $Likes_i$ and $Comments_i$.

We employ the DiD model in Equation (5) to identify the impact of virtual influencers on the number of sponsoring brands (for influencers) and the number of sponsored influencers (for brands) under the parallel trends assumption.

$$\phi^{DiD} = \frac{1}{n}\sum_{i=1}^{n} Y_{i1} - \frac{1}{n}\sum_{i=1}^{n} Y_{i0} \tag{5}$$

$$Y_{i1} = Y_1\left(\overrightarrow{X_{it}}\right) = \hat{Y}_1\left(\overrightarrow{X_{it}}\right) + \epsilon_{it}^{1}, Y_{i0} = Y_0\left(\overrightarrow{X_{it}}\right) = \hat{Y}_0\left(\overrightarrow{X_{it}}\right) + \epsilon_{it}^{0}$$

IPW Estimator

In Equation (6), we use the propensity score for treatment to quantify the treatment effect using the IPW estimator.

$$\phi^{IPW} = \frac{1}{n}\sum_{i=1}^{n} \frac{Z_{it}Y_{it}}{\hat{e}\left(\overrightarrow{X_{it}}\right)} - \frac{1}{n}\sum_{i=1}^{n} \frac{(1 - Z_{it})Y_{it}}{1 - \hat{e}\left(\overrightarrow{X_{it}}\right)} \tag{6}$$

$$\hat{e}\left(\overrightarrow{X_{it}}\right) = P\left(Z_{it} = 1 \middle| \overrightarrow{X_{it}}\right)$$

4.3. Heterogeneous Treatment Effect (HTE) Identification

We use HTE models to investigate whether *human influencers are differentially hurt by the introduction of virtual influencers based on demographics, appearance, or tenure* and whether *brand decisions vary systematically by product category*. We investigate sources of heterogeneity that are identified in the literature as characteristics that influence employees' vulnerability to AI displacement: age, gender, and race (Arntz et al., 2017; Colombo et al., 2019) as well as facial attractiveness and whether the person looks younger or older than their chronological age (Hosoda et al., 2003; Kaufmann et al., 2016). We include tenure even though Colombo et al. (2019) found that years of work experience did not affect one's vulnerability to displacement. Finally, at the brand level, we test for differences between brands in experience goods categories and brands in all other categories because experience goods are inherently more related to humanness than non-experience goods (Luo et al., 2019; Longoni et al., 2019).

To enhance interpretability, we use effect coding for the moderators following Datta, Knox, and Bronnenberg (2018). For example, $Attractive_{it} = 1$ for an above-median score on facial attractiveness, and $Attractive_{it} = 0$ for a below-median score. We use Equation (7) to estimate the HTEs ($\Gamma^h, \widetilde{\Gamma^h}$ on the influencer's outcomes (Y_{it}). We include vectors for demographics and appearance as well as dummies for tenure and product category (experience goods or not).

$$Y_{it} = \gamma_0^h + \gamma_1^h Z_i + \gamma_2^h Z_t + Z_{it} * \overrightarrow{H_{it}}' * \Gamma^h + g\left(\overrightarrow{X_{it}}\right) + \epsilon_{it} \tag{7}$$

$$\overrightarrow{H_{it}} = \begin{Bmatrix} \overrightarrow{Demographics_{it}^h}\ (Section\ 5.2.1), \overrightarrow{Appearance_{it}^h}\ (Section\ 5.2.2), \\ Tenure_{it}^h (Section\ 5.2.2), Experience\_good_{it}^h (Section\ 5.2.3) \end{Bmatrix}$$

$$\overrightarrow{Demographics_{it}^h} = \left[Age_{it}^h, Male_{it}^h, Race_{it}^h\right]'$$
$$\overrightarrow{Appearance_{it}^h} = \left[Attractive_{it}^h, Younger\_look_{it}^h\right]'$$

4.4. Mechanism Exploration

Brands naturally wish to hire influencers who will achieve the most engagement on sponsored posts, as higher engagement leads to larger marketing profits. Thus, we examine engagement on posts to gain insight into the mechanisms underlying the ATE and HTEs.

4.4.1. ATE Mechanism

Consumers increasingly are questioning the authenticity of human influencers who accept sponsorships (Hwang et al., 2021). Skeptical consumers may be less likely to engage, so influencers may struggle to get as much engagement on sponsored posts as they normally receive on non-sponsored posts. Brands might prefer virtual influencers if they do not suffer from the same authenticity problem as human influencers, such that virtual influencers can achieve comparable engagement on sponsored and non-sponsored posts. We test this possibility by regressing consumer engagement on the binary sponsorship identifier, using post-level data from both human and virtual influencers. We expect that the virtual influencers will achieve similar or higher engagement than human influencers overall ($\mu_1 \geq 0$), and although both human and virtual influencers will receive less engagement on sponsored posts than on non-sponsored posts ($\mu_2 < 0$, $\mu_4 < 0$), we expect sponsorship disclosure to be *less* damaging for the virtual influencers . Note that we run two models for a robustness check: without influencer fixed effects in Equation (8) and with influencer fixed effects in Equation (9). (Thus, Equation (9) does not contain fixed effects for the virtual influencers.)

$$\log\left(1+\frac{Engagements_{ijt}}{Followers_{ijt}}\right) = \mu_0 + \mu_1 * Virtual_i + \mu_2 * Sponsored_{it} + \mu_3 * Treat_i + \mu_4 * Treat_{it} + \tau_t + \epsilon_{ijt} \tag{8}$$

$$\log\left(1+\frac{Engagements_{ijt}}{Followers_{ijt}}\right) = \widetilde{\mu_0} + \widetilde{\mu_1} * Sponsored_{it} + \widetilde{\mu_2} * Treat_{it} + I_i + \tau_t + \epsilon_{ijt} \tag{9}$$

where $Engagements_{ijt}$ is the sum of likes and comments on post *j*, and $Followers_{ijt}$ is the number of followers of influencer *i* when *j* was posted at time *t*. Following the industry standard, we use the log-transformed number of engagements divided by the number of followers.[36]

4.4.2. HTE Mechanism: Influencer Demographics, Appearance, and Tenure

Human influencers may be subject to sources of heterogeneity that do not apply to virtual influencers, or vice versa. For example, if consumers tend to engage more with posts by younger (vs. older) human influencers but do not differentiate among virtual influencers on the basis of age, then brands might decide to replace some of their older human influencers with virtual

[36] This can be called "Instagram user engagement per post.".

influencers. To test whether engagement might explain some of the HTEs on the basis of influencer demographics, appearance, or tenure, we estimate Equation (10):

$$Engagements^{h}_{ijt} = \eta^{h}_{0} + \eta^{h}_{1} * \overrightarrow{H_{\iota t}}' * M^{h} + \epsilon^{h}_{ijt} \tag{10}$$
$$Engagements^{v}_{ijt} = \eta^{v}_{0} + \eta^{v}_{1} * \overrightarrow{V_{\iota t}}' * M^{v} + \epsilon^{v}_{ijt}$$

where

$$\overrightarrow{H_{\iota t}} = \left[Male^{H}_{i}, \log(Age^{H}_{it}), \overrightarrow{Race^{H}_{\iota}}, \log(Attractive^{H}_{i}), Elapsed\_days^{H}_{it}, Older\_looking^{H}_{i}\right]'$$

$$\overrightarrow{V_{\iota t}} = \begin{bmatrix} Male^{V}_{i}, \log(PredictedAge^{V}_{it}), \overrightarrow{PredictedRace^{V}_{\iota}}, \log(Attractive^{V}_{i}), \\ Elapsed\_days^{V}_{it}, Older\_looking^{V}_{i} \end{bmatrix}'$$

where $Predicted\_age^{v}_{it}$, and $Predicted\_race^{v}_{it}$ are the predicted age and race of virtual influencers, and the other variables are as defined elsewhere. We compare the coefficients of the human influencers and virtual influencers to understand how heterogeneity in engagement on the basis of influencer characteristics differs between human influencers and virtual influencers.

4.4.3. HTE Mechanism: Product Category (Experience Goods)

Finally, we estimate Equation (11) on the subpopulation of virtual influencers who posted sponsored content for experience goods brands. We conjecture that virtual influencers are not as effective at earning consumer trust (which we proxy as engagement) when they advertise experience goods, which have inherently human qualities. Thus, we expect either a negative or insignificant (but not positive) coefficient of $Experience\_goods_{ijt}$:

$$Engagements^{v}_{ijt} = \widetilde{\eta^{v}_{0}} + \widetilde{\eta^{v}_{1}} * Sponsored_{ijt} + \widetilde{\eta^{v}_{2}} * Experience\_goods_{ijt} + \epsilon^{v}_{ijt} \tag{11}$$

4.5. Changes in Verb Usage in Response to Treatment

Human influencers may adapt their marketing strategies to secure sponsorship deals when threatened by the rise of virtual influencers, analogous to the soft-selling strategy reported by Hwang et al. (2021). Specifically, we investigate changes in verb usage. We expect to find that treated influencers increased their usage of verbs that reflect cognitive processes, social skills, and senses that pertain to humans but not to AI (Colombo et al., 2019).

We extract all verbs used by all influencers in our sample, and we use the VerbNet hierarchy (Schuler 2005) to quantify the appearance of each verb type. VerbNet contains syntactic and semantic information about 233 verb types, such as verbs related to helping, future having, and consuming. We regress the number of appearances of each of the 233 verb types on the treatment and interaction variables:

$$\log\left(1 + Number\_of\_appearances_{imt}\right) = \eta_0 + \eta_1 Z_t + \eta_2 Z_{it} + g\left(\overrightarrow{X_{it}}\right) + I_i + \epsilon_{imt} \quad (12)$$

where $Number\_of\_appearance_{imt}$ is the log-transformed number of appearances of verb type m in posts by influencer *i* at time *t*.

5. Results

Section 5.1 reports the ATEs at the influencer level and brand level as estimated by the DR-DiD, DiD, and IPW estimators. Section 5.2 presents the HTEs by influencer demographics, appearance variables, and tenure as well as by the brand's product category (experience goods or not). Section 5.3 describes the results of our investigation into the possible mechanisms of the ATE and HTEs. Section 5.4 evaluates whether human influencers responded to the threat posed by virtual influencers by changing their verb usage.

5.1. Average Treatment Effects (ATE)

5.1.1. ATEs on Sponsorship Deals for Influencers & by Brands

To rigorously answer the question of whether the introduction of virtual influencers affected the sponsorship outcomes of human influencers, we examine (1) whether treated influencers, relative to control influencers, obtained sponsorship deals from more or fewer brands in the post-treatment period (influencer-level; supply side) and (2) whether treated brands, relative to control brands, hired more or fewer influencers in the post-treatment period (brand-level; demand side). The estimated ATEs from the DR-DiD estimator are in Table 7, column (1). We find that treated influencers suffered a 16.81% (=exp(-0.184)-1) decrease in the number of sponsorships in the post-treatment periods. This implies that virtual influencers displaced human influencers, such that brands that chose to partner with virtual influencers reduced the number of sponsorships provided to the human influencers with whom the brands already had relationships (i.e., the treated influencers). Surprisingly, however, treated brands *increased* their number of sponsorships after treatment, as shown in Table 7, column (2): treated brands increased the number of sponsored influencers by 18.53% (=exp(0.170)-1).

Table 7. ATE Estimations from the DR-DiD Estimator

	Estimates (Std. Err.)			
	(1) DV: Number of Brands (Influencer-Level Model)		**(2) DV: Number of Influencers (Brand-Level Model)**	
ATE	**-0.184***	**0.031**	**0.170***	**(0.078)**
Confidence Interval	(-0.245, -0.124)		(0.017, 0.323)	
Notes. All DVs are logged. * indicates that the confidence interval does not include 0. The ATE values are the aggregated effects from the yearly ATE estimates. As the DR-DiD estimation requires an equal number of periods for all treated and control units, we impute each late comer's values as zero (e.g., an influencer who joined in 2013 has zero sponsoring brands in 2011 and 2012).				

Figure 8 displays the yearly ATE effects estimated by DR-DiD (Callaway and Sant'Anna, 2021). The red bar lines indicate the pre-treatment periods, while the blue bar lines indicate the post-treatment periods. At the influencer level, the ATE becomes increasingly negative as the post-treatment period progresses; at the brand level, the ATE becomes increasingly positive.

Figure 8. Evolution in Yearly ATE

Influencer-Level ATE Yearly Trend Plot	Brand-Level ATE Yearly Trend Plot

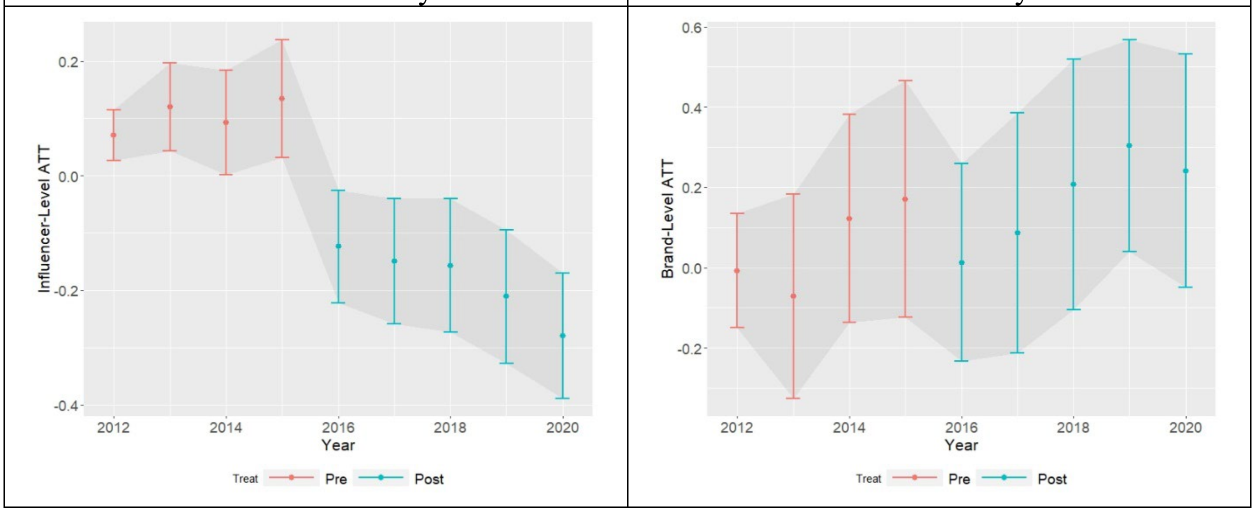


Table 8 presents the DiD and IPW estimation results. The results are qualitatively consistent with those in Table 7: treated influencers suffered a decrease in sponsorships of 5.92% (=exp(-0.061)-1) according to the DiD estimate and of 8.7% (=exp(-0.091)-1) according to the IPW estimate, while treated brands sponsored 25.35% more influencers (=exp(0.226)-1) according to the DiD estimate and 20.56% more influencers (=exp(0.187)-1) according to the IPW estimate. We conclude that the main results are robust to the model specification.

Why is the ATE positive for treated brands but negative for treated influencers? That is, how did employment growth at the brand level (the number of influencers sponsored) turn into an apparent reduction in employment at the influencer level (the number of sponsoring brands)? We reconcile the seemingly contradictory findings in the following section.

Table 8. ATE Estimations from the DiD and IPW Estimators

	Influencer-Level		**Brand-Level**	
	DiD	**IPW**	**DiD**	**IPW**
$\boldsymbol{Treat_{it}}$ (**$\boldsymbol{Treat_{kt}}$ for brand**)	**-0.061****	**-0.091*****	**0.226*****	**0.187****
	(0.022)	**(0.026)**	**(0.054)**	**(0.070)**
Observation	9,741	9,741	1,980	1,980
Influencer (brand) FE	Yes	No	Yes	No
Time FE	Yes	Yes	Yes	Yes
Control Vars	Yes	Yes	Yes	Yes
R^2	0.599	0.199	0.685	0.306
Notes. The extended results, with the progressive addition of control variables, are reported in Online Appendix 4.B. All DVs are logged. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001				

5.1.2. Reconciling the Discrepancy Between Influencer- and Brand-Level ATE Results

We hypothesize that treated influencers may have lost sponsorship opportunities because treated brands hired more control influencers (alongside virtual influencers) rather than treated influencers. We test this hypothesis with the DiD influencer-level models with three dependent variables: (1) the number of brands that sponsored treated influencers, (2) the number of brands that sponsored control influencers, and (3) the proportion of sponsoring brands that hired treated (rather than control) influencers. Likewise, we run three brand-level models: (4) the number of influencers sponsored by treated brands, (5) the number of influencers sponsored by control brands, and (6) the proportion of sponsored influencers who were hired by treated (rather than control) brands.

We report the results in Table 9. We found a 26.01% decrease (=exp(-0.302)-1) in the number of brands that sponsored treated influencers and a 17.55% decrease in the proportion of sponsoring brands that hired treated (rather than control) influencers (=exp(-0.193)-1). Meanwhile, we found a 10.63% increase (=exp(0.101)-1) in the number of brands that sponsored control influencers. The brand-level findings are similar: the number of influencers sponsored by treated brands did not change, the proportion of sponsorships from treated brands decreased, and the number of influencers sponsored by control brands soared by 107.09% (=exp(0.728)-1). This indicates that virtual influencers *displaced* treated human influencers but *complemented* control human influencers, validating both the AI displacement and complementarity effects for different subsets of the influencer population (Acemoglu and Restrepo, 2018).

Table 9. Influencers (inf) & Brand DiD Model Results

	(1) DV: # brands that sponsor treated infs	(2) DV: # brands that sponsor control infs	(3) DV: % brands that sponsor treated infs
$Treated\_inf_i$	0.501***	-0.001	0.291***
	(0.012)	(0.021)	(0.007)
$\boldsymbol{Treat\_inf_i \times Post_t}$	**-0.302*** **	**0.101*** **	**-0.193*** **
	(0.016)	**(0.026)**	**(0.010)**
Observation	9,741	9,741	6,809
Fixed Effects	Year & month, influencer demographics, appearance, content topics		
R^2	0.207	0.169	0.19
	(4) DV: # infs sponsored by treated brands	**(5) DV: # infs sponsored by control brands**	**(6) DV: % infs sponsored by treated brands**
$Treated\_brand_k$	0.498***	-0.738***	0.312***
	(0.051)	(0.036)	(0.016)
$\boldsymbol{Treated\_brand_k \times Post_t}$	**-0.101**	**0.728*** **	**-0.179*** **
	(0.063)	**(0.049)**	**(0.021)**
Observation	1,980	1,980	1,576
Fixed Effects	Year & month, productivity, followers, engagement, posts per day, parent corporation, SIC code		
R^2	0.385	0.299	0.334
Notes. All DVs are logged. Extended results are in Online Appendix 4.C. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001			

5.2. Heterogeneous Treatment Effects (HTEs)

Having established the ATEs, we move on to investigate how the treatment effect varies with the influencer's (1) demographics, (2) appearance, (3) tenure as well as with the brand's (4) product category (experience goods or not).

5.2.1. Influencer Demographics (Age, Gender, Race)

Table 10 provides the estimated coefficients of the interaction effects between $Treat_{it}$ and three demographic dummies: $Older_{it}$, $Male_{i}$, and $Black_{i}$. Consistent with the prediction of Arntz et al. (2017), older and male influencers were more at risk of displacement from virtual influencers. Specifically, older treated influencers lost 9.43% (=exp(-0.099)-1) more sponsorships than their younger counterparts, and male treated influencers lost 6.48% (=exp(-0.067)-1) more sponsorships than their female counterparts. Black treated influencers did not lose more sponsorships than treated influencers of other races.

Table 10. Influencer-Level HTE Results: Demographics

	Estimates (Std. Err.)					
Variable	**(1) Number of Sponsoring Brands**		**(2) Number of Sponsoring Brands**		**(3) Number of Sponsoring Brands**	
$Treated\_inf_i$	0.290***	(0.020)	0.290***	(0.020)	0.289***	(0.020)
$\boldsymbol{Treat_{it}}$	**-0.036**	**(0.030)**	**-0.063***	**(0.029)**	**-0.093***	**(0.027)**
$\times$ $\boldsymbol{Older_{it}}$	**-0.099***	**(0.027)**				
$\times$ $\boldsymbol{Male_{i}}$			**-0.067***	**(0.029)**		
$\times$ $\boldsymbol{Black_{i}}$					**0.072**	**(0.045)**
Observation	9,741		9,741		9,741	
R^2	0.206		0.205		0.204	
Notes. All DVs are logged. *Older* is a dummy variable that equals 1 (0) if the chronological age of influencer *i* at time *t* was above (below) the median age in the sample. *Male* equals 1 (0) if influencer *i* is male (female). *Black* equals 1 (0) if influencer *i* is Black (any of the other five race categories). Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001						

5.2.2. Influencer Appearance (Attractiveness, Older-Looking) and Tenure

Table 11 provides the estimated coefficients of the interaction effects between $Treat_{it}$ and $Attractive_{i}$, $Older\_look_{i}$, and $Short\_tenure_{it}$. We find an attractiveness premium, consistent with the literature (Hosoda et al., 2003): more-attractive treated influencers suffered a decrease in sponsorships of only 18.78% (=exp(-0.839+0.631)-1), while their less-attractive counterparts suffered a decrease of 56.79% (=exp(-0.839)-1). Surprisingly, we find that influencers who appear older than their chronological age fared better than influencers who appear younger than their age—older-looking treated influencers suffered a decrease in sponsorships of only 2.08%

(=exp(-0.115+0.094)-1), while their younger-appearing counterparts suffered a decrease of 10.86% (=exp(-0.115)-1). The result contradicts the prior finding that older-looking candidates are perceived as less healthy and are less likely to be hired (Kaufmann et al., 2016). Finally, tenure did not moderate the treatment effect, consistent with Colombo et al. (2019).

Table 11. Influencer-Level HTE Results: Appearance and Tenure

	Estimates (Std. Err.)					
Variable	**(1) Number of Sponsoring Brands**		**(2) Number of Sponsoring Brands**		**(3) Number of Sponsoring Brands**	
$Treated\_inf_i$	0.291***	(0.020)	0.290***	(0.020)	0.289***	(0.020)
$\boldsymbol{Treat_{it}}$	**-0.839***	**(0.173)**	**-0.115***	**(0.027)**	**-0.090**	**(0.027)**
$\times$ $\boldsymbol{Attractive_i}$	**0.631***	**(0.144)**				
$\times$ $\boldsymbol{Older\_look_i}$			**0.094**	**(0.030)**		
$\times$ $\boldsymbol{Short\_tenure_{it}}$					**0.018**	**(0.033)**
Observation	9,741		9,741		9,741	
R^2	0.206		0.205		0.204	
Notes. All DVs are logged. *Attractive* is a dummy variable that equals 1 (0) if influencer *i*'s facial attractive score is above (below) the sample median. *Older_look* equals 1 (0) if influencer *i*'s predicted age is greater (less than) their chronological age. *Short_tenure* equals 1 (0) if influencer *i*'s years of experience on Instagram at time *t* was below (above) the sample median. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001						

5.2.3. Product Category (Experience Goods)

Table 12 provides the estimated coefficients of the interaction effect between $Treat_{it}$ and $Experience_{good_k}$.[37] As we predicted, treated brands in the experience goods category increased the number of sponsored human influencers more than treated brands in other product categories in the post-treatment period. Our results are consistent with prior findings that people are reluctant to embrace AI in domains that seem especially human (Luo et al., 2019; Longoni et al., 2019). Perhaps aware of consumers' aversion to AI, brands in the experience goods category continued to rely heavily on human influencers even when virtual influencers were available.

[37] We consider 11 brands to be in the experience goods category: Airbnb, Ebay, Equinox, Netflix, Patreon, Plated, Spotify, Statefarm, Supercuts, Tinder, and Uber.

Table 12. Brand-Level HTE Results: Product Category

	DV: Number of Influencers Sponsored	
Variable	Coef.	SE
$Treated\_brand_k$	0.030	(0.063)
$\boldsymbol{Treat_{kt}}$	0.184*	(0.080)
$\boldsymbol{Treat_{kt} \times Experience\_good_k}$	0.848***	(0.128)
Observation	1,980	
FE	Year & month, productivity, followers, engagements, posts per day, parent corporation, SIC code	
R^2	0.309	
Notes. All DVs are logged. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001.		

5.3. Mechanisms of ATE & HTEs

5.3.1. Evidence of ATE Mechanism

In Table 13, columns (1) and (2) provide the results from estimating Equations (8) and (9); the results are comparable with and without influencer fixed effects, so we focus on the results in column (1) here. We find that virtual influencers received more engagement (normalized by the number of followers) than treated human influencers (coefficient of *Virtual* = 0.020). More importantly, while treated human influencers got 0.007 *less* engagement on sponsored posts than on non-sponsored posts (coefficient of *Sponsored* = -0.007), virtual influencers got *more* engagement (coefficient of *Virtual* + coefficient of *Sponsored* = 0.020 - 0.007 = 0.013). From the brand's perspective, engagement on posts is critical for the profitability of sponsorships, so the results help explain why brands found virtual influencers more attractive than treated human influencers for sponsorships.

Table 13. ATE Mechanism: Post Engagement by Sponsorship and Influencer Type

	Estimates (Std. Err.)			
	(1) DV: Engagement Divided by the Number of Followers		**(2) DV: Engagement Divided by the Number of Followers**	
$\boldsymbol{Virtual_i}$	**0.020*** **	**(0.001)**	**-**	**-**
$\boldsymbol{Treated_i}$	**0.004*** **	**(0.0003)**	**-**	**-**

$\boldsymbol{Sponsored_{ij}}$	**-0.007*** **	**(0.0002)**	**-0.002*** **	**(0.0004)**
$\boldsymbol{Treated_i \times Post_i}$	**-0.007*** **	**(0.0003)**	**-0.005*** **	**(0.001)**
Observation	118,976		118,976	
Individual FE (# of Influencers)	No		Yes (669 treated, 669 control, 137 virtual)	
Year-Month FE	Yes		Yes	
R^2	0.050		0.345	
Notes. DVs are log-transformed post-level engagement (the sum of likes and comments) divided by the number of followers. Robust standard errors are in parentheses. The coefficients for the post-treatment period are incorporated into the year-month fixed effects. *p<0.05; **p<0.01; ***p<0.001.				

5.3.2. Evidence of Influencer-Level HTE Mechanism

Table 14 provides the results of estimating Equation (10). More-attractive virtual influencers received more engagement, but we find no significant heterogeneity in engagement with virtual influencers by gender, age, race, or tenure. Meanwhile, human influencers have considerable heterogeneity: Engagement was significantly lower for male influencers, more-attractive influencers, influencers who are chronologically older, and influencers who appear older than their chronological age. Engagement is significantly higher for Black influencers and more experienced influencers (i.e., longer tenure).

Table 14. HTE Mechanism: Post Engagement by Demographics, Appearance, and Tenure

	(1) Human Inf. Post Engagement			**(2) Virtual Inf. Post Engagement**	
Variable	Coef.	SE	Variable	Coef.	SE
$\boldsymbol{Male_i}$	**-0.170*** **	**(0.025)**	$\boldsymbol{Male_i}$	**1.532**	**(2.400)**
$\mathbf{log}\,(\boldsymbol{Age_{ij}})$	**-1.490*** **	**(0.057)**	$\mathbf{log}(\boldsymbol{Age_{ij}})$	**7.966**	**(7.213)**
$\boldsymbol{Black_i}$	**0.171*** **	**(0.040)**	$\boldsymbol{Black_i}$	**-0.085**	**(0.049)**
$\mathbf{log}(\boldsymbol{Attractive_i})$	**-0.223***	**(0.105)**	$\mathbf{log}(\boldsymbol{Attractive_i})$	**10.765***	**(4.939)**
$\boldsymbol{Elapsed\_days_{ij}}$	**0.325*** **	**(0.035)**	$\boldsymbol{Elapsed\_days_{ij}}$	**-0.147**	**(0.352)**
$\boldsymbol{Older\_looking_i}$	**-0.111*** **	**(0.028)**			
Observation	31,225		Observation	953	
Year-Month FEs	Yes		Year-Month FEs	Yes	
Unreported Coefs	Races except for black		Unreported Coefs	Races except for black	
R^2	0.166		R^2	0.867	

Notes. DVs are log-transformed consumer engagement (total interactions; likes and comments). Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001.

5.3.3. Evidence of Brand-Level HTE Mechanism

Table 15 reports the results from estimating Equation (11) on the subsample of posts by virtual influencers with sponsorships in the experience goods category. Consistent with our conjecture, the coefficient of $Experience\_goods_j$ is not statistically significant, implying that virtual influencers do not get significantly more engagement on posts in the experience goods category than on posts in other categories.

Table 15. HTE Mechanism: Post Engagement with Virtual Influencers by Product Category

	DV: Virtual Influencers' Post Engagements	
Variable	Coef.	SE
$\boldsymbol{Sponsored_{ij}}$	**0.036**	**(0.032)**
$\boldsymbol{Experience\_goods_j}$	**-0.091**	**(0.112)**
Observation	31,603	
Year-Month Fixed Effects	Yes	
R^2	0.871	
Notes. The DV is the log-transformed consumer engagement (total interactions; likes and comments). Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001.		

5.4. Change in Verb Usage in the Post-Treatment Period

We estimated Equation (12) on all 233 verb types classified by VerbNet (Schuler 2005). Table 16 reports results from the seven estimations that yielded significant, positive treatment coefficients (full results are in Online Appendix 4.D),[38] meaning that usage increased among treated influencers in the post-treatment period. The seven verb types are related to helping, future having, consuming, growing, winking, stalking, and complaining. Table 17 lists the specific verbs in each of the seven types. As we expected, the seven verb types reflect

[38] We also found a significant result for verb type "berry," but we dropped the result because the words are nouns rather than verbs.

humanitarian behaviors (e.g., “help,” “cede”) as well as abilities that apply to humans but not machines (e.g., “spend,” “stalk,” “boast”). The results support our hypothesis that human influencers who are threatened with AI displacement may strategically change their posting behaviors to feature human-oriented activities, extending the research implications from Luo et al. (2019) and Longoni et al. (2019).

Table 16. Results for the Seven Verb Types with Significant Increases in Usage

	Estimates (Std. Err.)			
Variable	**(1) Help Verbs**	**(2) Future Having Verbs**	**(3) Consume Verbs**	**(4) Grow Verbs**
$\boldsymbol{Treated_i \times Post_t}$	**0.069***	**0.068***	**0.061***	**0.049***
	(0.029)	**(0.029)**	**(0.028)**	**(0.023)**
Observation	9,705	9,705	9,705	9,705
FE	Influencer, year/month, demographics, appearance, and engagement			
R^2	0.749	0.777	0.757	0.603
	Estimates (Std. Err.)			
	(5) Wink Verbs	**(6) Stalk Verbs**	**(7) Complain Verbs**	
$\boldsymbol{Treated_i \times Post_t}$	**0.045***	**0.044***	**0.029***	
	(0.023)	**(0.022)**	**(0.014)**	
Observation	9,705	9,705	9,705	
FE	Influencer, year/month, demographics, appearance, and engagement			
R^2	0.595	0.549	0.461	
Notes. The reported results are the models with coefficients of $\boldsymbol{Treated_i \times Post_t}$ that are significant at a level of 0.05 and positive. *p<0.05; **p<0.01; ***p<0.001.				

Table 17. The Seven Verb Types with Significant Increases in Usage

Verb Type	**Associated Verbs**
Help	abet • aid • assist • help • succor • support
Future having	allot • apportion • assign • award • bequeath • cede • extend • grant • offer • owe • portion • promise • ration • vote • will • yield
Consume	pass • spend • use
Grow	develop • grow • hatch
Wink	beckon • blink • clap • nod • point • shrug • squint • wag • wave • wink
Stalk	smell • stalk • track
Complain	boast • brag • complain • crab • gripe • grouch • kvetch • object

6. Robustness & Validity Checks

Online Appendix 5 reports the details of our robustness and validity checks. We show that the DiD results hold when the unit of analysis is by month (5.A). We replicate the results with the unmatched data in the DiD model (5.B) and DR-DiD model (5.C). Finally, we find no systematic or structured negative popularity shock (5.D), which increases our confidence that the estimated treatment effects are due to the introduction of virtual influencers rather than an external, negative popularity shock.

7. Conclusion

The rapid advance of AI in myriad domains is raising questions about whether AI workers tend to displace or complement the original human workers. We investigate the question empirically in an increasingly relevant labor market: social media influencers. We use a combination of deep representational learning algorithms and three causal inference models to characterize the average and heterogeneous effects of the introduction of virtual influencers to Instagram. We find that, among the brands that employed virtual influencers, human influencers who previously were sponsored by those brands tended to be displaced (16.81% fewer sponsorships in the post-treatment period), while human influencers who were *not* previously sponsored by those brands tended to be complemented (10.63% more sponsorships in the post-treatment period). The results support both a displacement effect (Acemoglu and Restrepo, 2018; Brynjolfsson et al., 2018; Grace et al., 2018; Webb, 2020) and a complementarity effect (Arntz et al., 2017; Felten et al., 2018; Colombo et al., 2019).

The human influencers varied considerably in their vulnerability to displacement. The most vulnerable were male influencers, older influencers, and less-attractive influencers, all consistent with the literature (Arntz et al., 2017; Hosoda et al., 2003; Kaufmann et al., 2016). The influencer's tenure (i.e., years on Instagram) did not moderate the treatment effect, consistent with Colombo et al. (2019). Surprisingly, and contradicting Kaufmann et al. (2016), influencers who appeared older than their chronological age were *less* vulnerable to displacement than their younger-appearing counterparts. We also found heterogeneity at the brand level: brands in the experience goods category increased their sponsorship of human influencers more in the post-treatment period than brands in other product categories. The result suggests that human influencers may remain competitive with virtual influencers to the extent that brands need to market the humanness of their offerings.

Finally, our results unveil a potential coping strategy used by human influencers who are threatened by the introduction of virtual influencers: an increase in the usage of human-specialized or human-oriented verbs. Although we did not test whether the strategy is *effective* at attracting engagement, sponsorships, or other desirable outcomes for influencers, the significant shift in posting behavior among treated influencers provides evidence of a new Luddite movement to avoid job displacement and showcase the importance of human workers in the influencer market.

Our identification of novel empirical effects of AI disruption in the job market of social media influencers has practical implications for governments, firms, and influencers. Our identification of the AI displacement effect and the most at-risk subpopulations may help the government prepare education or training for human influencers. By specifying that brands in the experience goods industry have not changed their sponsorship decisions as much as other brands, we provide insights for firms on how human-specialized businesses hire influencers. Lastly, we warn human influencers that virtual influencers are becoming an attractive marketing channel, and we reveal one strategy with which human influencers already seem to be combating virtual influencers (though we cannot comment on the effectiveness of the strategy). For all market practitioners, AI displacement may be closer than it appears.

8. References

Acemoglu, D., & Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6), 1488-1542.

Acemoglu, D., Lelarge, C., & Restrepo, P. (2020, May). Competing with robots: Firm-level evidence from France. In AEA Papers and Proceedings (Vol. 110, pp. 383-88).

Arntz, M., Gregory, T., & Zierahn, U. (2017). Revisiting the risk of automation. *Economics Letters*, 159, 157-160.

Athey, S., & Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89(1), 133-161.

Baker, M. J., & Churchill Jr, G. A. (1977). The impact of physically attractive models on advertising evaluations. *Journal of Marketing research*, 14(4), 538-555.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 3, 993-1022.

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy? *In AEA Papers and Proceedings* (Vol. 108, pp. 43-47).

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200-230.

Colombo, E., Mercorio, F., & Mezzanzanica, M. (2019). AI meets labor market: Exploring the link between automation and skills. *Information Economics and Policy*, 47, 27-37.

Datta, H., Knox, G., & Bronnenberg, B. J. (2018). Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science*, 37(1), 5-21.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Felten, E. W., Raj, M., & Seamans, R. (2018, May). A method to link advances in artificial intelligence to occupational abilities. *In AEA papers and proceedings* (Vol. 108, pp. 54-57).

Fildes R., & Goodwin P. (2007) Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, 37(6):570–576.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change*, 114, 254-280.

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2017). When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research*, 62, 729-754.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. *Personnel psychology*, 56(2), 431-462.

Hwang, S., Liu, X., & Srinivasan, K. (2021). Voice Analytics of Online Influencers—Soft Selling in Branded Videos. Available at SSRN: https://ssrn.com/abstract=3773825

Imbens, G. W. (2020). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *Journal of Economic Literature*, 58(4), 1129-79.

Kaufmann, M. C., Krings, F., & Sczesny, S. (2016). Looking too old? How an older age appearance reduces chances of being hired. *British journal of management*, 27(4), 727-739.
Li, F. (2019). Double-robust estimation in difference-in-differences with an application to traffic safety evaluation. Observational Studies, 5(1), 1-23.
Liang, L., Lin, L., Jin, L., Xie, D., & Li, M. (2018). SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. *In 2018 24th International Conference on Pattern Recognition* (ICPR) (pp. 1598-1603). IEEE.
Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4), 629-650.
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.
Ong, Thuy (2020). Lil Miquela, Lol's Seraphine: Virtual influencers make more real money than ever. *Bloomberg Businessweek*. https://www.bloomberg.com/news/features/2020-10-29/lil-miquela-lol-s-seraphine-virtual-influencers-make-more-real-money-than-ever
Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688.
Sanders, N. R., & Manrodt K. B. (2003) The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega*, 31(6):511–522.
Schuler, K. K. (2005). VerbNet: A broad-coverage, comprehensive verb lexicon.
Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., & Demirbas, M. (2010). Short text classification in Twitter to improve information filtering. *In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 841-842).
Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1701-1708).
Tang, Z., Zhang, X., Huang, L., & Dai, Y. (2013). Robust image hashing using ring-based entropies. *Signal Processing*, 93(7), 2061-2069.
Webb, Michael (2020). The Impact of Artificial Intelligence on the Labor Market, *Working paper*.

METAVERSE IS NEAR: THE IMPACT OF VIRTUAL INFLUENCERS

9. Online Appendix

Table of Contents

OA SECTION 1. EXAMPLE OF VIRTUAL INFLUENCER SPONSORSHIP

Figure S1 is an example of a sponsored post for TurboTax from a virtual influencer, Colonel Sanders, for KFC.

Figure S1. Example of a Sponsored Post from a Virtual Influencer

OA SECTION 2. DATA COLLECTION PROCESS

We construct a panel dataset of Instagram posts from virtual and human influencers from 2011 to 2020. In April 2020, we used virtualhumans.org, a company that publishes news, research, interviews, and biographies about virtual influencers, to collect the names of all 136 active virtual influencers. We focus on the 52 virtual influencers, listed in Table S1, that were sponsored by at least one brand.

Table S1. List of the 52 Virtual Influencers That Were Sponsored as of 2020

No.	Instagram Account Name	No.	Instagram Account Name
1	ai_angelica	27	lilmiquela
2	amara_gram	28	magazineluiza
3	amiyamato	29	mayaaa.gram
4	astrolovesu	30	mikuhatsune
5	barbie	31	milla_sakurai

6	bee_nfluencer	32	noonoouri
7	blawko22	33	phoenixmcewan
8	bodybyralph	34	poka_pokaka
9	boffothebear	35	pol.songs
10	brenn.gram	36	polishboy08
11	dagny.gram	37	reahkeem
12	dayzeeandstaxx	38	realqaiqai
13	esther.olofsson	39	ria_ria_tokyo
14	frenchgaia	40	robinabree
15	galaxia.gram	41	ruby.economics
16	gflaserbolt	42	ruby9100m
17	guggimon	43	ryan.seoul.icon
18	imma.gram	44	scazy
19	iongottlich	45	serahreikka
20	itsminaswrld	46	shudu.gram
21	janky	47	soymar.ia
22	jedyvales	48	thalasya_
23	kimzulu_	49	totinos
24	knoxfrost	50	world_record_egg
25	leyalovenature	51	yoox
26	liam_nikuro	52	zeevaah

From the posts of virtual influencers, we identify 287 sponsoring brands (listed in Table S2). Then, we identify brands that existed in the market before 2016 and sponsored human influencers (our *treated brands*; indicated with “Y” in the “Treated?” column of Table S2), and we match them with 130 *control brands* based on the brand characteristics used in Acemoglu et al. (2020).

Table S2. List of 287 Brands that Sponsored Virtual Influencers, and Whether They Also Sponsored Human Influencers Before 2016

No.	Brand Name	Treated?	No.	Brand Name	Treated?
1	Absolutvodka	Y	145	Madam Figaro Fr	Y
2	Acoldwall		146	Magnum	Y
3	Adidas	Y	147	Marc Jacobs	Y
4	Afi Sa		148	Marc Jacobs Fragrances	Y
5	Airbnb	Y	149	Marianne Fassler	
6	Azzedinealaia	Y	150	Mariano Vivianco	Y
7	Alberta Ferreti	Y	151	Masterclass	
8	Alex And Revauthier	Y	152	Max Mara	Y
9	Alexachung		153	Maybelline	Y
10	Alexander Mcqueen		154	Mbfwrussia	Y
11	Ali Express		155	MCM Worldwide	Y
12	Alibaba		156	Mert And Marcus	Y
13	Amazon		157	Mgllmn	

14	Amfar	Y	158	Mini Vision Urbanaut	Y
15	Anna Dello Russo	Y	159	Misbhv	Y
16	Anna Sui		160	Miss Shop	
17	Annevest		161	Missoni	Y
18	Aouadi		162	Miu Miu	Y
19	Ardene		163	Moncler	Y
20	Art Basel	Y	164	Moschino	Y
21	Balenciaga	Y	165	Msgm	
22	Barbie	Y	166	Mrs Hilfiger	Y
23	Beehome.net		167	Mugler	
24	Benq		168	Nadeshot	
25	Berliner Bags		169	Napanion | Planting Trees	
26	Bershka		170	Nativeinstruments	Y
27	Bogani Costantino		171	Nelly.com	
28	Bornxraised		172	Neocha	
29	Boss (Hugo Boss)		173	Netflix	Y
30	Brewdog Outpost Rotterdam		174	Off White	
31	Buddy Help		175	Onitsuka Tiger	
32	Bulgari	Y	176	Openingceremony	Y
33	Burberry	Y	177	Originalfunko	
34	Burt's Bees	Y	178	Oscarmayer	Y
35	Calvin Klein	Y	179	Osklen	Y
36	Carine Roitfeld	Y	180	Parallel Space Hk	
37	Cartier	Y	181	Patou	
38	Casa Do Rio		182	Patreon	Y
39	Celine		183	Peptalk	
40	Chanel	Y	184	Perrier	
41	Chiaraferragni	Y	185	Philosophy	Y
42	Chocolatos Id		186	Piaget	
43	Christian Louboutin (louboutinworld)	Y	187	Pierre Cardin	
44	Citeo		188	Pierre Herme	
45	Cniluxury		189	Pinko	Y
46	Coach	Y	190	Plus1org	
47	Collusion Studios		191	Popbee	
48	Colt45everytime		192	Porsche	Y
49	Cosmopolitan (cosmoindia)	Y	193	Postillion Hotels	
50	Cosmopolitan (cosmopolitan)	Y	194	Prada	Y
51	Covergirl	Y	195	Pro Bikegear	
52	Crate And Barrel		196	Puma	Y
53	Crayola	Y	197	Purell	
54	Cryptonmedia		198	Radboudnsm	
55	David Off		199	Ralph And Russo	Y
56	Dazed		200	Real Valentino Fashion	Y
57	Del Taco		201	Ramirez	
58	Dickies		202	Red Bull	Y

59	Dior	Y	203	Reebok	Y
60	Dolly Parton	Y	204	Reebok Classics	Y
61	Dundasworld		205	Rhizomedotorg	
62	Dvf	Y	206	Rich Mnisi	
63	EE Limited		207	Richard Leeds Int	
64	Eleven Paris		208	Ricola France	
65	Elle	Y	209	Roberto Cavalli	Y
66	Ellesse		210	Roger Vivier	Y
67	Emrata		211	Rola	
68	Emilio Pucci	Y	212	Rose Bergdorf	
69	Equinox	Y	213	Rotterdamtopsport	
70	Escada	Y	214	Sacai	Y
71	Etam		215	Sainthoax	
72	Etro	Y	216	Sako7	
73	Eurosport		217	Samsung Mobile	Y
74	Faena Hotel		218	Samsung UK	Y
75	Fashion Trust Arabia		219	Sandals Resorts Uk	
76	Fashion4relief		220	Sander Bos	
77	Fendi	Y	221	Schiaparelli	
78	Fenty		222	Scoob	
79	Ferragamo	Y	223	Segeraretreat	
80	Fleurs D'ici		224	Sheshouldrun	
81	Floss		225	Siemens	
82	Fondation De France		226	Skii	Y
83	Formula 1		227	Skims	
84	Frankie Morello		228	Smart Worldwide	
85	Furla		229	Stella Mccartney	Y
86	Galaxia		230	Stephen Jones Milinery	Y
87	Galeries Lafayette		231	Stonewall	
88	Genny		232	Stuart Weitzman	Y
89	Ghd Hair	Y	233	Sugarfina	Y
90	Giambattistavalli		234	Supreme	Y
91	Givenchy	Y	235	Swarovski	Y
92	Glamour Brazil	Y	236	The Marc Jacobs	Y
93	Gofundme		237	Team Vitality	
94	Gucci	Y	238	The Middle House	
95	Guerlain	Y	239	The Ouai	
96	Gunze		240	The Pioneer Woman	Y
97	Happy Egg		241	Thebe Magugu	
98	Hartmagazine		242	Thomas J Hilfiger	Y
99	Headspace		243	Tinder	Y
100	Hellmann's Mexico		244	Tods	Y
101	Hello Kitty	Y	245	Toga	
102	Heron Preston	Y	246	Tom Ford	Y
103	Highsnobiety	Y	247	Tommy Hilfiger	Y
104	H&M	Y	248	Tomo Koizumi	

105	Hypebeast	Y	249	Toysrus	Y
106	I Am Africa		250	Tramando	
107	Ikea		251	Trevor Stuurman	
108	Illy Coffee		252	Trojan Brand Condoms	
109	Img Sims Worldwide		253	Tropicofc	
110	Irisvanherpen	Y	254	Trussardi	
111	Isabel Marant	Y	255	Tynker Coding	
112	Istituto Marangoni		256	Uber	Y
113	Jacquemus		257	Ubereats	
114	Jeremy Scott	Y	258	Unicef	Y
115	Jil Sander		259	Unicef USA	Y
116	Jlo	Y	260	Unilever	Y
117	Jmd Helmets		261	V Magazine	Y
118	John John Denim	Y	262	Valentino	Y
119	Jpgaultier	Y	263	Veet Sa	
120	Juan De La Paz		264	Velo	
121	Karl Lagerfeld	Y	265	Verdy	
122	Karmagawa		266	Versace	Y
123	Kfc	Y	267	Vetements	
124	Kikaku		268	Vice Magazine	
125	Kingnature Ag		269	Victoria Beckham	Y
126	Kiss Tokyo Paper		270	Viktor And Rolf	
127	Kitstokickcancer		271	Vivianne Westwood	Y
128	Knorr		272	Vogue Australia	Y
129	La Kritza Clo		273	Vogue Brasil	Y
130	La Poste		274	Vogue Italia	Y
131	Lacoste	Y	275	Vogue Magazine	Y
132	Laduma		276	Worldaidsday	
133	Lancaster Beauty		277	Wwd	Y
134	Lancome	Y	278	Yeezy	
135	Lanvin	Y	279	Yoon Ambush	
136	Le Lis Blanc		280	Yoox	Y
137	Le Royal Monceau		281	Youporn	
138	Leica	Y	282	Youtube Music	
139	Les Caves Lechapelais		283	YouTube	
140	Lexus		284	Yves Saint Laurent	Y
141	Lofficiel Italia	Y	285	Zattini	
142	Loreal Paris		286	Zimmermann	
143	Love Philosophy	Y	287	Zu Hair Murad	Y
144	Louis Vuitton	Y			

Our *treated influencers* are the 669 human influencers who were sponsored by the 130 treated brands ***before*** 2016, identified with Instagram's official tool for tracking sponsorship relationships and ten kinds of sponsorship disclosure language, summarized in Table S3 (based on the post data from before 2016).

Then, we identify eligible control influencers: those who used a treated brand's products or include a treated brand's name in their Instagram content (but did not have any sponsored posts with a treated brand before 2016). We construct a sample of 669 *control influencers* by matching them with the treated influencers using a propensity score based on demographics, appearance, and post characteristics.

Table S3. Common Types of Sponsorship Disclosure Language

No.	Sponsorship Disclosure Language	Proportion	Similar Languages
1.	Thank you "Brand" for sponsoring	72.5%	Thanks "Brand" for sponsoring
2.	#ad	13.5%	#ad, #adv, #advertisement
3.	A "Brand" ambassador	1.7%	I become a virtual ambassador of "Brand."
4.	Collaborated with "Brand"	1.5%	Collaboration (collab) with "Brand"
5.	Partnered with "Brand"	1.1%	Partnered up (partnership) with "Brand"
6.	Sponsored by "Brand"	<1%	"Brand" is sponsoring today's post
7.	Teamed up with "Brand"	<1%	Teaming up with "Brand"
8.	Supported by "Brand"	<1%	"Brand" supports this post
9.	Powered by "Brand"	<1%	This post is empowered by "Brand"
10.	An advertisement for "Brand"	<1%	This content is an advertisement of "Brand"

Notes. The first column lists the most representative form of the sponsorship disclosure language with distinctive root words. The second column shows the proportion of sponsored posts in our sample (from before 2016) that contained the language. The third column lists derivatives in the same category of disclosure language.

OA SECTION 3. CONTROL VARIABLE REPRESENTATIONS & POST MATCHING

OA 3.A. MODEL FOR AGE, GENDER, RACE, AND YOUNGER- VS. OLDER-LOOKING

DeepFace is a deep neural network architecture that predicts demographic information such as age, gender, and race based on a person's photo. We use Multi-Task Cascaded Convolutional Networks (MTCNN) to detect the face of the influencer in each picture.

For human influencers, we use DeepFace to predict race (six options: Asian, Black, Indian, Latino/Hispanic, Middle Eastern, or White) and age.[39] We compare each influencer's predicted age with

[39] For human influencers, we collected chronological age, gender, and birth country from the third-party influencer profile website, famousbirthday.com.

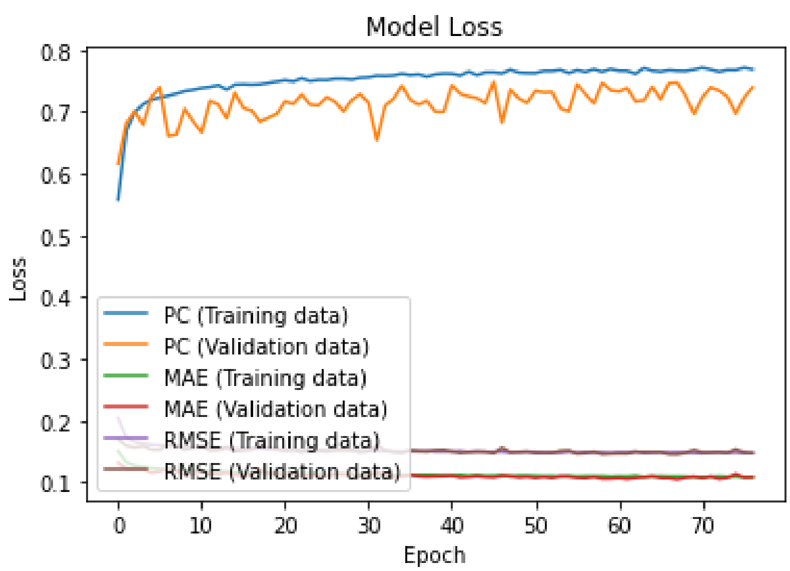
their chronological age to determine whether they are “younger-looking” or “older-looking.” For instance, if an influencer is 25 years old, but DeepFace predicts an age of 20, then we label her as younger-looking.

For virtual influencers, we use DeepFace to predict age, gender, and race.

OA 3.B: MODEL FOR ATTRACTIVENESS

We start with ResNet-50, a customizable CNN, and we add ImageNet weights and five more dense layers to improve the benchmark model performance. Our ResNet-55 model has a Mean Absolute Error (MAE) loss value of 0.1087 and a Rooted Mean Squared Error (RMSE) loss value of 0.1477, which compare favorably with the performance of the similar model in Liang et al. (2018) (MAE = 0.2291, RMSE = 0.3017). We report the loss value convergence plot for the ResNet-55 model in Figure S2. The Pearson correlation coefficient (PC), MAE, and RMSE converge after 80 epochs.

Figure S2. Loss Value Convergence Plot for the Resnet-55 Model: PC, MAE, and RMSE



OA 3.C: MODEL FOR CONTENT TOPICS

Figure S3 shows the coherence scores of models with 2–15 topics; the 11-topic model has the best score, 0.5112 (Blei et al., 2003; Newman et al., 2010). Figure S4 shows word clouds for each of the 11 topics, and Figure S5 shows the distribution of the topics in posts. The most prevalent is Topic 6, which contains fashion-related word stems such as “fashion,” “shoe,” and “bag.” The least prevalent is Topic 5, which is related to word stems such as “mem” and “today.”

Figure S3. Coherence Score Distribution by the Number of Topics (2–15)

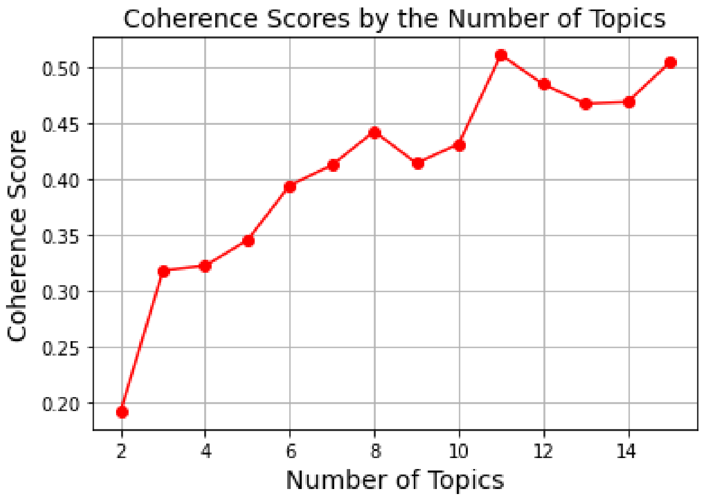
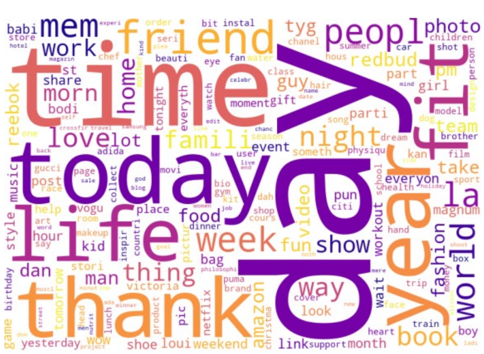
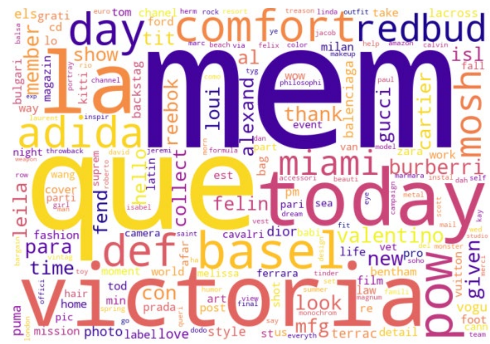


Figure S4. Word Clouds of the 11 Topics in the LDA-BERT Autoencoder Model

(Topic 0)	(Topic 1)
	
(Topic 2)	(Topic 3)
	
(Topic 4)	(Topic 5)



(Topic 6)

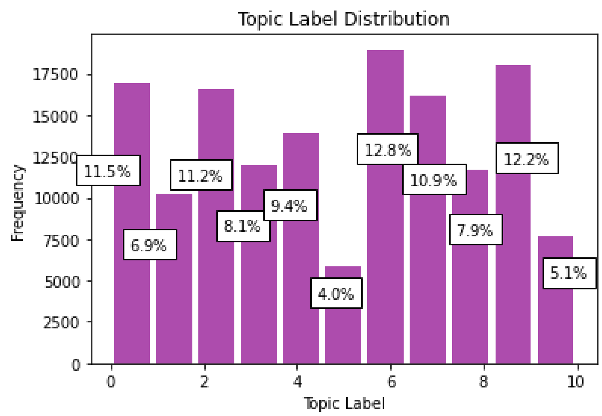
(Topic 7)

(Topic 8)

(Topic 9)

(Topic 10)

Figure S5. Topic Label Distribution



OA 3.D: AIC/BIC COMPARISON OF PSM MODEL SPECIFICATIONS

We test several PSM model specifications (Caliendo and Sabine, 2008), and Table S4 reports the performances. For influencer-level matching, we achieve the best performance when we include demographics (age, gender, and race), appearance (attractiveness and younger- vs. older-looking), tenure, content topics, and other post characteristics (total posts, sponsored posts, followers, likes, and comments). For brand-level matching, we achieve the best performance when we include the SIC code, large firm dummies, average productivity, and sponsored post characteristics (average number of posts per day, followers, likes, and comments) in the pre-treatment period.

Table S4. AIC & BIC Results of PSM Model Specifications

Influencer-Level Matching Model Specifications	AIC	BIC
Model 1a) Demographics (Demo) + Appearance (Appr) + Tenure	2463.092	2524.138
Model 2a) Demo + Appr + Tenure + Topics	2377.465	2494.007
Model 3a) Demo + Appr + Tenure + Topics + Monthly total posts, sponsored posts, followers, likes, and comments in the pre-treatment period	2348.211	4090.789
Model 4a) Demo + Appr + Tenure + Topics + Last month total posts, sponsored posts, followers, likes, and comments in the pre-treatment period (2015-12)	2325.565	2469.855
Model 5a) Demo + Appr + Tenure + Topics + Average total posts, sponsored posts, followers, likes, and comments in the pre-treatment period (proposed)	**2163.433**	**2307.723**
Brand-Level Matching Model Specifications	AIC	BIC
Model 1b) SIC + Large firm dummies + Monthly productivity, posts per day, followers, likes, and comments in the pre-treatment period	6466.985	8203.438
Model 2b) SIC + Large firm dummies	826.246	1282.685
Model 3b) SIC + Large firm dummies + Last month's productivity, posts per day, followers, likes, and comments in the pre-treatment period (2015-12)	815.65	1301.857
Model 4b) SIC + Large firm dummies + Average productivity, posts per day, followers, likes, and comments in the pre-treatment period (proposed)	**813.912**	**1300.119**

OA SECTION 4. DID, IPW & DR-DID ESTIMATIONS

The parameter of interest is the average treatment effect (ATE) on the treated (ATT). The major limitation of the DiD estimator, the dominant causal inference model, is that the data must satisfy the parallel trends assumption (Rosenbaum and Rudin 1983; Rosenbaum 2002).[40] Therefore, we use DR-DiD and IPW as well as DiD to ensure the robustness of the results. The estimators share the same aim: to identify the ATE in the equation below.

$$\phi = E\left[Y_{i,t=1} - Y_{i,t=0} \middle| D_i = 1\right] = E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 1]$$
$$= \frac{1}{N}\sum_{i|D=1}^{N} Y_{i1} - \frac{1}{N}\sum_{i|D=1}^{N} Y_{i0}$$

We estimate the treatment effect (ϕ) between the pre- and post-treatment periods with each identification strategy (ϕ^{DiD}, ϕ^{IPW} and , ϕ^{DR-DiD}). Also, we employ ϕ^{DiD}, ϕ^{IPW} in an alternative setting by year to estimate how the treatment effect evolves over time.

Comparison of the DiD, IPW, and DR-DiD Estimators

- DiD Estimator

$$\phi^{DiD} = \bar{Y}_{1,1} - \left[\bar{Y}_{1,0} + \frac{1}{n_{treat}} \sum_{i \mid D_i=1} \left(\hat{\mu}_{0,1}(\overrightarrow{X_i}) - \hat{\mu}_{0,0}(\overrightarrow{X_i})\right)\right]$$

where $\bar{Y}_{d,t} = \sum_{i|D_i=d,T_i=t} Y_{it}/n_{d,t}$,

$\bar{Y}_{d,t}$ is the sample average outcome among units in treatment group *d* at time *t,*

and

$\hat{\mu}_{d,t}(\overrightarrow{x_i})$ is the estimator of the true, unknown $m_{d,t}(\overrightarrow{x_i}) = E[Y_t|D = d, \overrightarrow{X_i} = \overrightarrow{x_i}]$

- IPW Estimator

$$\phi^{IPW} = \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{D_i - \hat{\pi}_i(\overrightarrow{X_i})}{1 - \hat{\pi}_i(\overrightarrow{X_i})}(Y_1 - Y_0)}{\frac{1}{n}\sum_{i=1}^{n} D_i}$$

$\hat{\pi}(\overrightarrow{X_i})$ is the estimator of the true, unknown propensity score $p(\overrightarrow{X_i})$

[40] We present the parallel trend plot in Section 6.2.2 in the main paper.

- DR-DiD Estimator

$$\phi^{DR-DiD} = E\left[\left(\omega_1(Z_i) - \omega_0\left(Z_i, \vec{X_i}; \pi\right)\right)\left(\Delta Y_i - \mu_{0,\Delta}\left(\vec{X_i}\right)\right)\right]$$

$$\omega_1(Z_i) = \frac{Z_i}{E[Z_i]}, \qquad \omega_0\left(Z_i, \vec{X_i}; g\right) = \frac{\frac{g(\vec{X_i})(1 - Z_i)}{1 - g(\vec{X_i})}}{E[\frac{g(\vec{X_i})(1 - Z_i)}{1 - g(\vec{X_i})}]}$$

OA 4.A. DR-DID ESTIMATION PROCESS

We explain the four-step process below as it applies to a setting with only two periods, pre-treatment and post-treatment. (The DR-DiD estimation for multiple periods is illustrated in Callaway and Sant'Anna, 2021.)

$$\phi^{DR-DiD} = E\left[\left(\omega_1(Z_i) - \omega_0\left(Z_i, \vec{X_i}; \pi\right)\right)\left(\Delta Y_i - \mu_{0,\Delta}\left(\vec{X_i}\right)\right)\right]$$

$$= \overset{4}{\boxed{E\left[\overset{2}{\boxed{\omega_1(Z_i)}} \times \left(\Delta Y_i - \overset{1}{\boxed{\mu_{0,\Delta}\left(\vec{X_i}\right)}}\right) - \overset{3}{\boxed{\omega_0\left(Z_i, \vec{X_i}; \pi\right)}} \times \left(\Delta Y_i - \overset{1}{\boxed{\mu_{0,\Delta}\left(\vec{X_i}\right)}}\right)\right]}}$$

$$\omega_1(Z_i) = \frac{Z_i}{E[Z_i]}, \qquad \omega_0\left(Z_i, \vec{X_i}; g\right) = \frac{\frac{g(\vec{X_i})(1 - Z_i)}{1 - g(\vec{X_i})}}{E[\frac{g(\vec{X_i})(1 - Z_i)}{1 - g(\vec{X_i})}]}$$

(4) Difference in Outcome Regression Fitted Value $\left(\mu^p_{0,\Delta}(X_i)\right)$

First, we compute the predicted value of the outcome regression for the control influencers, $\mu_{0,\Delta}\left(\vec{X_i}\right)$, using the following outcome regression model.

$$\Delta Y_0 = Y_{0,1} - Y_{0,0} = \mu_{0,\Delta}\left(\vec{X_i}\right) + \epsilon_i = \beta_0 + \vec{X_i} * \vec{\beta}' + \xi_i$$

(2) Weight for Treated Influencers' Outcomes

Then, we compute the weight for the treated influencers' outcomes.

$$\Omega_1(Z_i) = Z_i * \left(\frac{1}{N}\sum_i^N Z_i\right)^{-1}$$

(3) Weight for Control Influencers' Outcomes

This step is similar to the previous, but the control influencers did not receive treatment, so we employ the propensity score of getting treatment $\left(g(\overrightarrow{X_\iota})\right)$ based on the control influencers' covariates.

$$\Omega_0\left(Z_i, \overrightarrow{X_\iota}; \pi\right) = \frac{g(\overrightarrow{X_\iota})(1 - Z_i)}{1 - g(\overrightarrow{X_\iota})} * \left(\frac{1}{N}\sum_{i}^{N} \frac{g(\overrightarrow{X_\iota})(1 - Z_i)}{1 - g(\overrightarrow{X_\iota})}\right)^{-1}$$

$$\text{where } g\left(\overrightarrow{X_\iota}\right) = \hat{g}\left(\overrightarrow{X_\iota}\right) + \epsilon_i = \alpha_0 + \overrightarrow{X_\iota}' * A + \epsilon_i$$

(4) Taking the Average Treatment Effect (ATE) on the Treated (ATT)

Finally, we take the average of the treatment effect after accounting for the $E[\mathrm{Y}_{\mathrm{i},0} | D_i = 1]$.

$$\phi^{DR-DiD} = \frac{1}{N}\sum_{i=1}^{N}\left[\omega_1(Z_i) \times \left(\Delta Y_i - \mu_{0,\Delta}(\overrightarrow{X_\iota})\right) - \omega_0\left(Z_i, \overrightarrow{X_\iota}; \pi\right) \times \left(\Delta Y_i - \mu_{0,\Delta}(\overrightarrow{X_\iota})\right)\right]$$

OA 4.B: DID & IPW RESULTS WITH PROGRESSIVE ADDITIONS OF CONTROLS

Tables S5 to S8 provide the extended results of the DiD and IPW models with the progressive addition of control variables. The main results are consistent with the DR-DiD model results across the different model specifications. In Tables S5 and S6 (influencer-level), we find that the introduction of virtual influencers caused treated influencers to lose sponsorships. In Tables S7 and S8 (brand-level), our results show that brands that sponsored virtual influencers hired even more human influencers in the post-treatment period.

Table S5. Influencer-Level ATE via DiD Estimator

	DV: # of Brands			
	(1) $\boldsymbol{Treat_{it}}$	**(2) $\boldsymbol{Treat_{it}}$ & Time FE**	**(3) $\boldsymbol{Treat_{it}}$ & Time FE & Controls**	**(4) $\boldsymbol{Treat_{it}}$ & User/Time FEs & Controls**
$\boldsymbol{Treated\_inf_i}$	**0.334***	**0.340***	**0.289***	
	(0.021)	**(0.021)**	**(0.020)**	
$\boldsymbol{Post_t}$	**-0.148***			
	(0.018)			
$\boldsymbol{Treated\_inf_i \times Post_t}$	**-0.098***	**-0.104***	**-0.086**	**-0.061**
	(0.028)	**(0.028)**	**(0.026)**	**(0.022)**
log(1+ Followers)			-0.024*	0.056**
			(0.009)	(0.018)
log(1+ Likes)			0.055***	0.025
			(0.012)	(0.016)
log(1+ Comments)			0.035*	0.037*

			(0.014)	(0.017)
log(1+ # of Posts)			0.249***	0.243***
			(0.010)	(0.014)
Constant	0.795***	0.626***	-0.218	-3.591***
	(0.013)	(0.122)	(0.264)	(0.815)
Observation	9,741	9,741	9,741	9,741
Individual FE	No	No	No	Yes
Time FE	No	Yes	Yes	Yes
Control Vars	No	No	Yes	Yes
R^2	0.054	0.079	0.204	0.599
Notes. For parsimony, we incorporate the coefficients and SEs of demographics, appearance variables, tenure, and content topics as controls. All DVs are logged. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001.				

Table S6. Influencer-Level ATE via IPW Estimator

	DV: # of Brands			
	(1) $Treat_{it}$	**(2) $Treat_{it}$ & Time FE**	**(3) $Treat_{it}$ & Time FE & Controls**	**(4) $Treat_{it}$ & User/Time FEs & Controls**
$Treated\_inf_i$	**0.334*** **	**0.340*** **	**0.288*** **	
	(0.021)	**(0.021)**	**(0.021)**	
$Post_t$	**-0.148*** **			
	(0.018)			
$Treated\_inf_i \times Post_t$	**-0.098*** **	**-0.104*** **	**-0.089*** **	**-0.061** **
	(0.028)	**(0.028)**	**(0.026)**	**(0.022)**
log(1+ Followers)			-0.034***	0.056**
			(0.010)	(0.018)
log(1+ Likes)			0.061***	0.025
			(0.014)	(0.016)
log(1+ Comments)			0.048**	0.037*
			(0.015)	(0.017)
log(1+ # of Posts)			0.264***	0.243***
			(0.011)	(0.014)
Constant	0.795***	0.626***	-0.24	-3.591***
	(0.013)	(0.122)	(0.277)	(0.815)
Observation	9,741	9,741	9,741	9,741
Individual FE	No	No	No	Yes
Time FE	No	Yes	Yes	Yes
Control Vars	No	No	Yes	No
R^2	0.04	0.071	0.201	0.599

Notes. For parsimony, we incorporate the coefficients and SEs of demographics, appearance variables, tenure, and content topics as controls. All DVs are logged. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001.

Table S7. Brand-Level ATE via DiD Estimator

	DV: # of Influencers			
	(1) $Treat_{kt}$	**(2) $Treat_{kt}$ & Time FE**	**(3) $Treat_{kt}$ & Time FE & Controls**	**(4) $Treat_{kt}$ & User/Time FEs & Controls**
$\boldsymbol{Treated\_brand_k}$	**0.205*** **	**0.192*** **	**0.055**	
	(0.058)	**(0.057)**	**(0.053)**	
$\boldsymbol{Post_t}$	**-0.130** **			
	(0.049)			
$\boldsymbol{Treated\_brand_k \times Post_t}$	**0.218** **	**0.231** **	**0.217** **	**0.202*** **
	(0.077)	**(0.076)**	**(0.066)**	**(0.053)**
log(1+Productivity)			0.087***	0.072***
			(0.010)	(0.013)
log(1+Followers)			0.052***	0.201***
			(0.009)	(0.038)
log(1+Likes)			-0.113***	0.051
			(0.019)	(0.037)
log(1+Comments)			0.194***	0.032
			(0.025)	(0.038)
log(1+Posts.Per.Day)			-0.033	0.119
			(0.053)	(0.070)
Constant	1.109***	0.693***	-0.373	-2.415***
	(0.036)	(0.00000)	(0.250)	(0.505)
Observation	1,980	1,980	1,980	1,980
Individual FE	No	No	No	Yes
Time FE	No	Yes	Yes	Yes
Control Vars	No	No	Yes	No
R^2	0.042	0.079	0.336	0.691
Notes. All DVs are logged. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001.				

Table S8. Brand-Level ATE via IPW Estimator

	DV: # of Brands			
	(1) $Treat_{kt}$	**(2) $Treat_{kt}$ & Time FE**	**(3) $Treat_{kt}$ & Time FE & Controls**	**(4) $Treat_{kt}$ & User/Time FEs & Controls**
$\boldsymbol{Treated\_brand_k}$	**0.205*** **	**0.192*** **	**0.068**	

	(0.058)	**(0.057)**	**(0.054)**	
$\boldsymbol{Post_t}$	**-0.130****			
	(0.049)			
$\boldsymbol{Treated\_brand_k \times Post_t}$	**0.218****	**0.231****	**0.188****	**0.202*****
	(0.077)	**(0.076)**	**(0.067)**	**(0.053)**
log(1+Productivity)			0.107***	0.072***
			(0.012)	(0.013)
log(1+Followers)			0.048***	0.201***
			(0.010)	(0.038)
log(1+Likes)			-0.122***	0.051
			(0.020)	(0.037)
log(1+Comments)			0.197***	0.032
			(0.027)	(0.038)
log(1+Posts.Per.Day)			-0.018	0.119
			(0.058)	(0.070)
Constant	1.109***	0.693***	-0.549	-2.415***
	(0.036)	(0.00000)	(0.310)	(0.505)
Observation	1,980	1,980	1,980	1,980
Individual FE	No	No	No	Yes
Time FE	No	Yes	Yes	Yes
Control Vars	No	No	Yes	No
R^2	0.039	0.084	0.336	0.691
Notes. All DVs are logged. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001.				

OA 4.C: EXTENDED RESULTS OF TABLE 9

Tables S9 and S10 display the extended results of the influencer-level DiD model. We estimate the model on four dependent variables: the number of brands that sponsor (1) influencers, (2) treated influencers, and (3) control influencers, as well as (4) the proportion of sponsoring brands that hired treated (rather than control) influencers. For a robustness check, we run the models with and without influencer fixed effects.

Tables S11 and S12 are analogous but pertain to the brand-level DiD model. Again, there are four dependent variables: the number of influencers sponsored by (5) brands, (6) treated brands, and (7) control brands as well as (8) the proportion of sponsored influencers who were sponsored by treated (rather than control) brands. Again, the two tables report results with and without fixed effects.

Table S9. Influencer-Level ATE via DiD Estimator, Without Influencer FEs

	Estimates (Std. Err.)

Variable	(1) DV: # brands that sponsor infs	(2) DV: # brands that sponsor treated infs	(3) DV: # brands that sponsor control infs	(4) DV: % brands that sponsor treated infs
$\boldsymbol{Treated\_inf_i}$	**0.289*** **	**0.501*** **	**-0.001**	**0.291*** **
	(0.020)	**(0.012)**	**(0.021)**	**(0.007)**
$\boldsymbol{Treated\_inf_i \times Post_t}$	**-0.086** **	**-0.302*** **	**0.101*** **	**-0.193*** **
	(0.026)	**(0.016)**	**(0.026)**	**(0.010)**
Observation	9,741	9,741	9,741	6,809
Influencer FE	No	No	No	No
Time FE	Yes	Yes	Yes	Yes
Control Vars	Yes	Yes	Yes	Yes
R^2	0.204	0.207	0.169	0.19
Notes. All DVs are logged. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001.				

Table S10. Influencer-Level ATE via DiD Estimator, With Influencer FEs

	Estimates (Std. Err.)			
Variable	**(1) DV: # brands that sponsor infs**	**(2) DV: # brands that sponsor treated infs**	**(3) DV: # brands that sponsor control infs**	**(4) DV: % brands that sponsor treated infs**
$\boldsymbol{Treated\_inf_i \times Post_t}$	**-0.061** **	**-0.306*** **	**0.135*** **	**-0.195*** **
	(0.022)	**(0.015)**	**(0.022)**	**(0.010)**
Observation	9,741	9,741	9,741	6,809
Influencer FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Control Vars	Yes	Yes	Yes	Yes
R^2	0.599	0.509	0.571	0.448
Notes. All DVs are logged. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001.				

Table S11. Brand-Level ATE via DiD Estimator, Without Brand FEs

	Estimates (Std. Err.)			
Variable	**(1) DV: # infs sponsored by brands**	**(2) DV: # infs sponsored by treated brands**	**(3) DV: # infs sponsored by control brands**	**(4) DV: % infs sponsored by treated brands**
$\boldsymbol{Treated\_brand_k}$	**0.055**	**0.498*** **	**-0.738*** **	**0.312*** **
	(0.053)	**(0.051)**	**(0.036)**	**(0.016)**
$\boldsymbol{Treated\_brand_k \times Post_t}$	**0.217** **	**-0.101**	**0.728*** **	**-0.179*** **
	(0.066)	**(0.063)**	**(0.049)**	**(0.021)**

Observation	1,980	1,980	1,980	1,576
Brand FE	No	No	No	No
Time FE	Yes	Yes	Yes	Yes
Control Vars	Yes	Yes	Yes	Yes
R^2	0.336	0.385	0.299	0.334
Notes. All DVs are logged. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001.				

Table S12. Brand-Level ATE via DiD Estimator, With Brand FEs

	Estimates (Std. Err.)			
Variable	**(1) DV: # infs sponsored by brands**	**(2) DV: # infs sponsored by treated brands**	**(3) DV: # infs sponsored by control brands**	**(4) DV: % infs sponsored by treated brands**
$\boldsymbol{Treated\_brand_k \times Post_t}$	**0.202*** **	**-0.093**	**0.699*** **	**-0.176*** **
	(0.053)	**(0.050)**	**(0.047)**	**(0.021)**
Observation	1,980	1,980	1,980	1,576
Brand FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Control Vars	Yes	Yes	Yes	Yes
R^2	0.691	0.701	0.574	0.552
Notes. All DVs are logged. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001.				

OA 4.D: FULL VERB USAGE RESULTS

We estimate Equation (12) in the main text on each of the 233 verb types classified by VerbNet (Schuler 2005) and report the full results in Table S13. The highlighted rows are presented in the main paper as they have positive, significant coefficients (i.e., a significant increase in usage among treated influencers in the post-treatment period). The results support our hypothesis that human influencers who are threatened with AI displacement may strategically change their posting behaviors to feature human-oriented activities.

Table S13. Full Results of Estimating Equation (12)

DV	Coef.	SE	P-Value	R^2
rush	-0.03165	0.010817	0.003443	0.394443
peer	-0.06857	0.024206	0.004623	0.805236
rummage	-0.06235	0.025102	0.013013	0.791854
amuse	-0.06046	0.02483	0.014915	0.805549
help	0.068989	0.028823	0.01671	0.748563

future_having	0.067815	0.028503	0.017373	0.776803
keep	-0.06481	0.027283	0.017557	0.714955
send	-0.06227	0.027011	0.021161	0.767035
confess	-0.03715	0.01692	0.028135	0.494255
consume	0.061285	0.028144	0.029465	0.756637
grow	0.049317	0.022776	0.030392	0.602759
refrain	-0.00652	0.003041	0.032006	0.183265
reflexive_appearance	-0.05637	0.026472	0.03324	0.753462
berry	0.047358	0.023153	0.040847	0.63294
complain	0.029282	0.014339	0.041169	0.461037
fulfilling	-0.05233	0.025779	0.042385	0.571034
stalk	0.043578	0.021663	0.044294	0.548949
wink	0.045309	0.022801	0.046934	0.595195
transfer_mesg	-0.05191	0.026797	0.052739	0.774506
exceed	-0.04966	0.025786	0.054162	0.656402
cheat	0.049252	0.026583	0.063948	0.766511
instr_communication	0.050923	0.02753	0.064387	0.628211
cost	-0.04208	0.023379	0.071902	0.798976
pain	-0.03673	0.020831	0.077924	0.566198
floss	0.034158	0.019757	0.083873	0.541824
dedicate	0.024421	0.014339	0.088598	0.533846
flinch	-0.00803	0.004731	0.089609	0.287379
light_emission	0.041906	0.02477	0.09072	0.624545
tingle	0.038651	0.023205	0.095823	0.581151
leave	0.023344	0.014553	0.108741	0.375464
swat	0.031777	0.020164	0.115086	0.492228
breathe	-0.0262	0.016968	0.122591	0.483773
hunt	0.041943	0.027334	0.124952	0.754431
build	-0.0368	0.02409	0.126697	0.812597
free	-0.03945	0.025921	0.12809	0.632422
carve	0.040128	0.026395	0.128474	0.686951
concealment	-0.03646	0.024063	0.129812	0.553845
contribute	-0.03382	0.022793	0.137906	0.603945
urge	0.037104	0.025111	0.139561	0.6619
dine	-0.036	0.024689	0.144821	0.696317
appear	0.03386	0.023757	0.154123	0.800927
inquire	0.035239	0.025146	0.161137	0.665248
body_internal_motion	-0.03489	0.02492	0.161516	0.636823
knead	-0.03419	0.025533	0.180621	0.776033
pour	0.021609	0.016373	0.186948	0.530447
crane	-0.03166	0.024067	0.18841	0.804292

spank	-0.03116	0.024056	0.19529	0.562259
substance_emission	-0.03029	0.023969	0.206353	0.595236
obtain	0.035263	0.028195	0.211077	0.688179
stimulus_subject	-0.03085	0.02499	0.217051	0.807218
register	-0.03229	0.026563	0.22414	0.660561
pit	0.033673	0.027718	0.224469	0.727278
orphan	-0.01023	0.008455	0.226294	0.300183
sound_existence	0.02199	0.018521	0.235134	0.49581
adopt	-0.0152	0.013409	0.257095	0.487057
contiguous_location	0.0271	0.024031	0.25947	0.810527
calve	0.017022	0.015196	0.262675	0.451609
illustrate	0.030877	0.027593	0.263176	0.689465
bill	-0.02818	0.025474	0.268663	0.642507
disappearance	0.021575	0.019691	0.273244	0.540821
begin	-0.02817	0.026168	0.28171	0.725091
appeal	0.027665	0.025845	0.28447	0.691309
defend	0.025171	0.023534	0.284852	0.581588
engender	-0.01856	0.017363	0.285253	0.512742
marry	0.026404	0.024907	0.289133	0.61079
body_internal_states	-0.01489	0.014114	0.291439	0.391749
say	0.027146	0.026348	0.30292	0.751371
nonverbal_expression	0.025868	0.025113	0.303023	0.638071
assessment	-0.01756	0.017333	0.311137	0.525889
declare	0.026789	0.02661	0.314102	0.689908
fit	0.025283	0.025399	0.31955	0.815341
butter	-0.02525	0.025597	0.323966	0.797202
hold	-0.02283	0.02346	0.33046	0.573537
separate	0.025728	0.026437	0.330496	0.665345
roll	-0.02474	0.025942	0.340218	0.74765
exhale	0.000254	0.000268	0.34281	0.126214
animal_sounds	0.02395	0.026322	0.362903	0.691764
throw	-0.02333	0.025676	0.363475	0.752251
convert	-0.02046	0.023095	0.375682	0.847352
matter	-0.02018	0.0228	0.376166	0.603033
push	-0.02223	0.025163	0.376952	0.612599
indicate	-0.02279	0.025967	0.380057	0.766842
drive	0.021145	0.024124	0.380766	0.574295
meet	-0.02209	0.02525	0.381745	0.761454
wish	0.022344	0.025685	0.384374	0.794261
poke	0.016491	0.019059	0.386936	0.495356
gorge	0.023185	0.026902	0.388796	0.732915

linger	-0.00987	0.011672	0.397899	0.345564
simple_dressing	0.023652	0.027984	0.398039	0.738866
allow	0.016948	0.020364	0.405287	0.560211
break	-0.02088	0.025657	0.415712	0.667566
settle	0.00681	0.00841	0.418104	0.410371
try	-0.0117	0.01472	0.426776	0.431589
performance	0.019159	0.024433	0.432968	0.804474
cooperate	-0.02049	0.026166	0.433629	0.765487
clear	-0.01753	0.022572	0.437385	0.586487
pelt	0.013462	0.017353	0.437915	0.436985
slide	-0.01956	0.025264	0.438869	0.655328
care	-0.02031	0.026651	0.446112	0.708153
focus	0.016309	0.021506	0.448279	0.584315
feeding	0.013103	0.017432	0.452265	0.482105
spray	0.019486	0.025928	0.452332	0.751593
suspect	0.004169	0.005559	0.45334	0.254562
reach	-0.019	0.025515	0.456448	0.791658
meander	-0.01805	0.02426	0.456965	0.816353
consider	0.017622	0.023736	0.457875	0.840009
modes_of_being_with_motion	0.018603	0.025122	0.459021	0.644152
sound_emission	-0.01916	0.02629	0.466198	0.746597
pay	0.017244	0.02426	0.477219	0.608349
assuming_position	0.018064	0.025676	0.481734	0.712654
banish	0.00334	0.004828	0.489076	0.473405
eat	0.017319	0.025154	0.49115	0.691238
bend	0.006724	0.009917	0.497748	0.47739
complete	-0.01314	0.019493	0.500408	0.562359
exist	0.016485	0.024711	0.504723	0.789867
order	0.017004	0.025612	0.506775	0.814204
differ	-0.01647	0.024809	0.506857	0.635968
sight	-0.01644	0.024822	0.507735	0.765575
suffocate	0.005017	0.007685	0.513907	0.281677
chase	-0.01845	0.028284	0.514271	0.685319
dub	0.015578	0.024682	0.527943	0.80339
risk	-0.00929	0.014909	0.53327	0.441596
admire	0.012953	0.020876	0.534963	0.851138
marvel	-0.01652	0.026764	0.536983	0.795704
carry	-0.01544	0.025208	0.540315	0.678535
exchange	0.007791	0.013116	0.55251	0.385806
conspire	-0.01365	0.023045	0.553803	0.593608
stop	-0.01537	0.025974	0.553926	0.720962

other_cos	-0.01301	0.022221	0.558251	0.842176
dressing_well	0.014825	0.02584	0.566165	0.73766
neglect	-0.00775	0.013688	0.571455	0.435148
change_bodily_state	-0.00399	0.00707	0.572356	0.285614
bump	0.014105	0.025394	0.578609	0.648464
cling	0.002515	0.004533	0.579081	0.3774
equip	0.012392	0.022569	0.582979	0.632584
cope	-0.01	0.01832	0.585185	0.545536
mine	-0.00987	0.018563	0.594863	0.468542
manner_speaking	0.0139	0.026572	0.60091	0.715313
herd	0.014686	0.028109	0.601355	0.653078
mix	-0.01559	0.030491	0.609264	0.752753
being_dressed	0.004265	0.008419	0.612452	0.253814
admit	0.010539	0.021188	0.618907	0.571169
remove	-0.00997	0.020208	0.621738	0.549112
weather	0.012191	0.025545	0.633215	0.678523
curtsey	0.006434	0.013585	0.635789	0.384717
calibratable_cos	-0.01232	0.026896	0.646796	0.746445
see	0.011121	0.024798	0.653833	0.815692
bring	0.011221	0.025455	0.659356	0.775761
nonvehicle	0.010783	0.024883	0.664786	0.635686
entity_specific_cos	0.010293	0.023812	0.665562	0.632478
avoid	-0.00601	0.013906	0.665605	0.448309
cut	-0.01082	0.025233	0.668046	0.641604
approve	0.009189	0.021626	0.670938	0.573601
captain	-0.01105	0.026121	0.672358	0.72137
withdraw	0.003948	0.009337	0.67243	0.483238
get	-0.00966	0.022863	0.672732	0.847085
groom	0.003752	0.008997	0.676693	0.399247
rely	0.008507	0.020542	0.678803	0.518456
murder	0.009163	0.022765	0.687315	0.574024
judgement	-0.00974	0.024299	0.688646	0.776485
coil	-0.00941	0.02351	0.689044	0.590654
appoint	0.010672	0.026889	0.691461	0.767328
talk	-0.00959	0.02655	0.717953	0.698729
put_direction	-0.00869	0.024317	0.720922	0.625413
image_impression	0.009599	0.027005	0.722267	0.715364
smell_emission	0.005651	0.016486	0.731765	0.427338
create	0.009486	0.027843	0.733343	0.730502
swarm	0.008766	0.025781	0.733867	0.651446
snooze	0.007484	0.022349	0.737743	0.541618

split	-0.0081	0.024965	0.745689	0.764963
succeed	-0.00454	0.014121	0.74801	0.430977
debone	-0.00106	0.003315	0.748972	0.179116
disassemble	0.003252	0.01032	0.752657	0.340207
investigate	0.006925	0.022372	0.756909	0.539325
characterize	-0.00703	0.022918	0.758963	0.847797
transcribe	-0.00838	0.028251	0.76673	0.667512
force	-0.00778	0.026375	0.767953	0.787327
confine	0.006648	0.023107	0.773566	0.600847
spatial_configuration	-0.00722	0.025304	0.775413	0.755559
accompany	-0.00542	0.019053	0.776137	0.554538
want	0.007209	0.025454	0.777015	0.830311
hiccup	-0.00544	0.019938	0.785029	0.569563
waltz	0.002837	0.010633	0.78963	0.316067
dress	0.006925	0.026269	0.792064	0.732265
gobble	0.002049	0.008708	0.814014	0.324422
steal	-0.00587	0.025114	0.815138	0.781905
wipe_instr	-0.00488	0.021085	0.817035	0.54089
advise	0.004241	0.018709	0.820667	0.516055
occurrence	-0.00568	0.025283	0.82237	0.662784
amalgamate	-0.00601	0.027203	0.825182	0.708042
lodge	-0.00544	0.025384	0.830419	0.77261
touch	-0.00503	0.02386	0.832886	0.564623
devour	-0.00126	0.006015	0.833637	0.292974
turn	-0.00513	0.024508	0.834107	0.650979
lecture	0.005764	0.027786	0.835673	0.73268
fill	0.004968	0.024463	0.839076	0.789712
give	-0.00535	0.026703	0.841066	0.740564
wipe_manner	0.005551	0.027791	0.841698	0.754977
pocket	-0.00502	0.025588	0.844427	0.807411
scribble	0.005004	0.02609	0.847893	0.707881
correspond	0.004962	0.025876	0.847922	0.639728
preparing	0.004813	0.025309	0.849188	0.785246
weekend	0.004698	0.026121	0.85726	0.748649
destroy	-0.00273	0.015205	0.857275	0.450364
put	0.004401	0.025714	0.864117	0.749614
poison	-0.0046	0.026963	0.864657	0.684505
long	-0.00433	0.02598	0.867653	0.754579
entity_specific_modes_being	0.004024	0.026427	0.878981	0.698822
chew	-0.00348	0.024234	0.885668	0.626043
learn	0.004737	0.034525	0.890881	0.681174

hit	0.003706	0.02703	0.890963	0.731915
put_spatial	-0.00337	0.025678	0.895436	0.715766
chit_chat	0.002796	0.022356	0.900489	0.523688
tell	-0.00311	0.025128	0.901512	0.668534
discover	0.003333	0.027645	0.904039	0.722918
funnel	-0.0032	0.026751	0.904696	0.720598
conjecture	0.002898	0.025016	0.907782	0.818577
shake	-0.00289	0.027438	0.916218	0.724855
ferret	0.001483	0.014105	0.916243	0.443829
coloring	0.002656	0.026126	0.919029	0.704652
run	-0.00236	0.024738	0.924062	0.798849
tape	0.002861	0.030145	0.924386	0.749657
battle	-0.00238	0.026951	0.929749	0.67606
search	-0.00199	0.025907	0.938924	0.777439
escape	0.001711	0.022715	0.939972	0.827662
hurt	0.001935	0.026716	0.942266	0.746825
vehicle	-0.00168	0.025056	0.9464	0.648907
forbid	0.001376	0.025573	0.957086	0.659866
limit	-0.00093	0.020125	0.963244	0.498311
price	-0.00069	0.019618	0.972144	0.529428
cooking	0.000552	0.025078	0.98244	0.693361
masquerade	-0.00032	0.024181	0.989464	0.594711
braid	-0.00016	0.025709	0.994913	0.74562

OA SECTION 5. ROBUSTNESS CHECKS

OA 5.A: DID RESULTS WITH MONTHLY DATA

We repeat the DiD estimation with data at the month level. The estimated treatment effects are qualitatively consistent with the main results in Table 8.

Table S14. DiD Estimation at the Month Level

Variable	**DV: # brands that sponsor infs**		Variable	**DV: # infs sponsored by brands**	
	(1) Time FEs & Controls	**(2) Infs/Time FEs & Controls**		**(3) Time FEs & Controls**	**(4) Brands/Time FEs & Controls**
$Treat_i$	0.072***		$Treat_k$	0.044***	
	(0.004)			(0.010)	
$\boldsymbol{Treat_{it}}$	**-0.030*** **	**-0.019*** **	$\boldsymbol{Treat_{kt}}$	**0.030***	**0.071*** **

	(0.004)	**(0.004)**		**(0.012)**	**(0.012)**
log(1+ Followers)	-0.004**	0.015***	log(1+Productivity)	0.024***	0.021***
	(0.001)	(0.003)		(0.002)	(0.003)
log(1+ Likes)	0.011***	0.001	log(1+Followers)	0.012***	0.008***
	(0.002)	(0.002)		(0.001)	(0.001)
log(1+ Comments)	0.008***	0.006**	log(1+Likes)	-0.035***	-0.013***
	(0.002)	(0.002)		(0.003)	(0.003)
log(1+ # of Posts)	0.075***	0.078***	log(1+Comments)	0.065***	0.026***
	(0.001)	(0.002)		(0.004)	(0.005)
			log(1+Posts.Per.Day)	0.003	0.105***
				(0.008)	(0.012)
Observation	109,285	109,285	Observation	22,352	22,352
Influencer FE	No	Yes	Brand FE	No	Yes
Time FE	Yes	Yes	Time FE	Yes	Yes
Controls	Yes	Yes	Controls	Yes	Yes
R^2	0.072	0.233	R^2	0.147	0.325

Notes. All DVs are logged. Robust standard errors are in parentheses. In the influencer-level models, unreported controls are demographics, appearance, tenure, and content topics. In the brand-level models, unreported controls are the large corporation dummies and SIC code fixed effects. *p<0.05; **p<0.01; ***p<0.001.

OA 5.B: DID RESULTS WITH UNMATCHED DATA

We repeat the DiD estimation with unmatched data. The estimated treatment effects are qualitatively consistent with the main results in Table 8.

Table S15. DiD Estimation with Unmatched Data

Variable	**DV: # brands that sponsor infs**		Variable	**DV: # infs sponsored by brands**	
	(1) Time FEs & Controls	**(2) Infs/Time FEs & Controls**		**(3) Time FEs & Controls**	**(4) Brands/Time FEs & Controls**
$Treat_i$	0.328***		$Treat_k$	0.034	
	(0.018)			(0.042)	
$\boldsymbol{Treat_{it}}$	**-0.092*** **	**-0.065*** **	$\boldsymbol{Treat_{kt}}$	**0.270*** **	**0.257*** **
	(0.023)	(0.020)		(0.054)	(0.041)
log(1+ Followers)	-0.022**	0.043**	log(1+Productivity)	0.077***	0.050***
	(0.008)	(0.015)		(0.005)	(0.006)
log(1+ Likes)	0.057***	0.027*	log(1+Followers)	0.038***	0.254***

	(0.010)	(0.013)		(0.004)	(0.017)
log(1+ Comments)	0.030**	0.038**	log(1+Likes)	-0.080***	0.018
	(0.011)	(0.015)		(0.010)	(0.013)
log(1+ # of Posts)	0.218***	0.227***	log(1+Comments)	0.161***	0.050***
	(0.008)	(0.012)		(0.012)	(0.015)
			log(1+Posts.Per.Day)	-0.085**	0.058
				(0.026)	(0.037)
Observation	13,635	13,635	Observation	7,973	7,973
Influencer FE	No	Yes	Brand FE	No	Yes
Time FE	Yes	Yes	Time FE	Yes	Yes
Controls	Yes	Yes	Controls	Yes	Yes
R^2	0.213	0.585	R^2	0.298	0.625
Notes. All DVs are logged. Robust standard errors are in parentheses. In the influencer-level models, unreported controls are demographics, appearance, tenure, and content topics. In the brand-level models, unreported controls are the large corporation dummies and SIC code fixed effects. *p<0.05; **p<0.01; ***p<0.001.					

OA 5.C: DR-DID RESULTS WITH UNMATCHED DATA

We repeat the DR-DiD estimation with unmatched data at the month level. The estimated treatment effects are qualitatively consistent with the main results in Table 7.

(Note that we could not repeat the DR-DID estimation on month-level data, as we did with the DiD estimator, because there was insufficient variation in the number of sponsoring brands and sponsored influencers by month.)

Table S16. DR-DiD Estimation Results for the Unmatched Influencers and Brands

Variable	**(1) DV: Number of Brands (Influencer-Level Model)**		**(2) DV: Number of Influencers (Brand-Level Model)**	
ATE	**-0.233***	**0.029**	**0.152***	**(0.064)**
Confidence Interval	(-0.290, -0.175)		(0.026, 0.277)	
Notes. All DVs are logged. * indicates that the confidence interval does not include 0. The ATE values are the aggregated effects from the yearly ATE estimates. As the DR-DiD estimation requires an equal number of periods for all treated and control units, we impute each late comer's values as zero (e.g., an influencer who joined in 2013 has zero sponsoring brands in 2011 and 2012).				

OA 5.D: TESTING FOR A POSSIBLE NEGATIVE POPULARITY SHOCK

If a negative popularity shock occurred in 2016, and if the shock affected treated influencers more than control influencers, then sponsorships may have fallen among treated influencers because of the shock rather than because of the introduction of virtual influencers. To verify that our treated influencers did not face a negative popularity shock,[41] we regress the rate of change in engagement on the treatment group indicator, as written in Equation (S1):

$$\log\left(1 + \frac{Engagements_{it}}{Engagements_{i,t-1}}\right) = \omega_0 + \omega_1 * Treated\_inf_i * Post_t + \Omega' * [Influencer\ FEs, Time\ FEs, Content\ FEs] + \epsilon_{it} \tag{S1}$$

The results are in Table S17. Our parameter of interest, ω_1, is not statistically significant at a level of 0.05. We conclude that a systematic negative popularity shock cannot account for our estimated treatment effect.

Table S17. Testing for a Negative Popularity Shock

	Estimates (Std. Err.)	
Variable	**(1) Log-transformed Engagement Rate Change**	
$\boldsymbol{Treated\_inf_i \times Post_t}$	**-0.0088**	**(0.0145)**
Observation	9,732	
Influencer FEs	Yes (1,338 influencers)	
Time FEs	Yes	
Control FEs	Yes	
R-squared	0.3837	
Notes. *p<0.05; **p<0.01; ***p<0.001.		

REFERENCES

Acemoglu, D., Lelarge, C., & Restrepo, P. (2020, May). Competing with robots: Firm-level evidence from France. In AEA Papers and Proceedings (Vol. 110, pp. 383-88).

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), "Latent dirichlet allocation," The Journal of Machine Learning research, (3), 993-1022.

Caliendo, Marco, and Sabine Kopeinig (2008), "Some practical guidance for the implementation of propensity score matching," Journal of economic surveys, 22 (1), 31-72.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200-230.

[41] It is highly unlikely that a brand is reluctant to hire more than 500 treated influencer subset due to the brand's subjective evaluation. Instead, we incorporate consumer engagement (the sum of likes and comments), following the law of large numbers, to reflect the overall popularity of each influencer.

Newman, David, et al. (2010), "Automatic evaluation of topic coherence," Human language technologies: The 2010 annual conference of the association for computational linguistics.
Rosenbaum, Paul R (2002), "Overt bias in observational studies." Observational studies. Springer, New York, NY, 71-104.
Rosenbaum, Paul R., and Donald B. Rubin (1983), "The central role of the propensity score in observational studies for causal effects," Biometrika, 70 (1), 41-55.
Schuler, K. K. (2005). VerbNet: A broad-coverage, comprehensive verb lexicon.

3. Beyond Human: The Impacts of Human-Like Virtual Influencers on Consumer Engagement

ABSTRACT

Virtual influencers, characterized as digital personalities that exhibit human-like qualities on social media, have emerged as a prominent phenomenon in the marketing landscape. Prior research indicates that virtual influencers may provide novelty, but they could lack the authenticity and reliability that are commonly associated with human social influencers. Furthermore, there remains a paucity of understanding regarding how online audiences engage with and relate to these artificial digital creations. To fill the gap, we empirically test the impact of human-like qualities on consumer engagements using Instagram data. We use deep learning techniques to measure virtual influencer demographics (DeepFace for predicting age, gender and race, and emotions), image aesthetics (Neural Image Assessment), and human-like poses (Realtime Multi-Person Pose Estimation using Part Affinity Fields). Our findings indicate that virtual influencers can increase their comments and improve comment sentiments by increasing their interaction on Instagram comments. Additionally, writing more comments can lead to an increase in both new and revisiting comment writers. Aesthetically pleasing images and being female or having a certain ethnicity can also positively impact virtual influencers' engagement with commenting users. Surprisingly, new users are more likely to leave comments on virtual influencers with smaller followings, while revisiting users leave comments more frequently on familiar virtual influencers. Moreover, virtual influencers who exhibit emotions of fear are found to receive more consumer engagement from revisiting users. We provide practical implications for virtual influencer development firms, virtual influencers themselves, and the overall audience related to AI assistant development.

1. Introduction

Virtual influencers, characterized as digital personalities that exhibit human-like qualities on social media, have emerged as a prominent phenomenon in the marketing landscape. Notably, in fall 2018, the renowned French luxury fashion house Balmain unveiled a campaign featuring three digital models, including Shudu Gram, recognized as the world's first digital supermodel, who operates as a free agent and has gained significant popularity. This example illustrates the growing attention that virtual influencers have captured in the field of marketing. As AI-driven entities, virtual influencers have disrupted conventional notions of celebrity endorsement and brand promotion on social media platforms, challenging traditional marketing strategies and opening up new avenues for engaging consumers in novel and innovative ways.

While AI-driven CGI virtual influencer companies, such as Brud, are actively engaged in creating virtual influencers and formulating marketing strategies on Instagram, prior research indicates that virtual influencers may provide novelty but could lack the authenticity and reliability that are commonly associated with human social influencers (Moustakas et al., 2020). Furthermore, although there has been a growing interest within the research community regarding virtual influencers and their utilization in marketing strategies, there remains a paucity of understanding regarding how online audiences engage with and relate to these artificial digital creations (Stein et al., 2022).

To empirically understand how people perceive virtual influencers and their human-like qualities on social media, we incorporate Instagram post and comment data and quantify three different types of human qualities, namely (1) visual characteristics encompassing virtual influencers' predicted age, gender, race, emotions, and image aesthetic score, (2) estimated human-like poses, and (3) the frequency of comments written by virtual influencers on their own posts. Then we test how these human qualities affect consumer engagements on Instagram posts.

Specifically, we ask two research questions. First, how do virtual influencers' human-like qualities change the number of comments and the comment sentiment on the virtual influencers' Instagram posts? Second, how do the human-like qualities influence the growth of the new and revisiting (existing) users on Instagram comments?

We answer our research questions by leveraging deep learning methods to detect potential confounds such as attractiveness and content topics, and we construct matched sets of influencers and brands. We use a two-level identification strategy by comparing sponsorships received by treated vs. control influencers in the post-treatment period as well as sponsorships given by treated vs. control brands in the post-treatment period. We construct a doubly robust difference-in-differences (DR-DiD) estimator to avoid the model misspecification bias (Athey and Wager, 2017; Imbens, 2020; Callaway and Sant'Anna, 2021), and we confirm the findings with a difference-in-differences (DiD) estimator and inverse propensity weighting (IPW).

We find that virtual influencers are more likely to receive higher number of comments when their image aesthetics are perceived as more pleasing, they identify as female or Asian, and increase their participation in writing comments on their posts. Conversely, being identified as black is associated with receiving fewer comments. Additionally, we find that virtual influencers can consistently enhance comment sentiment by writing more comments on their own Instagram post. However, if influencers already have a substantial following, it may have a negative impact on comment sentiment.

Next, we also show that virtual influencers who possess a more aesthetic image, have Asian-oriented facial traits, and actively engage in writing Instagram post comments tend to receive higher levels of user engagement in the form of comments. Furthermore, we observe that new users are more likely to engage in leaving comments when there are higher levels of user engagement on the Instagram post image, when the influencer is female, and when the influencer has a smaller following. Conversely, revisiting users are more likely to leave comments when they perceive fear emotion on the influencer's face, when the influencer has a larger following, and when the influencer is not of black ethnicity. These results underscore the significance of virtual influencers actively writing comments to foster user engagement. Interestingly, we also find that user preferences regarding the popularity of virtual influencers differ, with new users showing a preference for smaller-sized influencers, and vice versa.

Our findings shed light on the complex and multifaceted nature of consumer perceptions of artificial intelligence (AI) and virtual influencers, including dimensions such as trust and perceived human-likeness. By elucidating these insights, our study provides valuable implications for marketers and practitioners seeking to effectively leverage AI-driven technologies, such as virtual influencers, in their marketing strategies.

2. Related Literature

The theoretical framework of our research draws upon and extends existing research on two key areas: (i) consumer perceptions of artificial intelligence (AI) and virtual influencers, and (ii) the effects of human-like qualities in virtual influencers on consumer behavior.

2.1. Consumer Trust of Artificial Intelligence (AI) and Virtual Influencers

One prominent aspect of consumer perceptions of AI is trust. Research has shown that consumers' trust in AI-driven technologies, such as virtual influencers, is influenced by factors such as transparency, explainability, reliability, and perceived competence of the AI system (Chen et al., 2019; Kim et al., 2020). For instance, studies have found that consumers are more likely to trust virtual influencers that are transparent about their AI-driven nature and provide explanations for their actions (Wang et al., 2018). Additionally, consumers tend to trust virtual influencers that are perceived as reliable and competent in delivering their intended messages (Choi et al., 2021).

2.2. The Effects of Human-like Qualities in Virtual Influencers on Consumer Behavior

Another important dimension of consumer perceptions of AI is perceived human-likeness. Consumers often form perceptions about the human-like qualities of virtual influencers, including their appearance, voice, behavior, and communication style. Research has shown that consumers tend to respond positively to virtual influencers that exhibit human-like traits, as they find them more relatable and authentic (Lu et al., 2019; Marwick & boyd, 2019). However, there is also evidence that consumers may have concerns about the authenticity and genuineness of virtual influencers, particularly if their human-likeness is perceived as misleading or deceptive (Gupta et al., 2020).

We make contributions to the existing literature in two key areas. Firstly, we empirically investigate consumer perceptions of virtual influencers' Instagram posts, answering how consumers perceive and respond to the content generated by virtual influencers. Secondly, we examine how human-like qualities of virtual influencers influence consumer engagements on Instagram, thereby providing insights into the differential effects of human-like virtual influence qualities on consumer behavior in the context of social media marketing. Our findings highlight that consumer perceptions of AI and virtual influencers are multifaceted, involving dimensions such as trust and perceived human-likeness. Understanding these consumer perceptions is essential for marketers and practitioners to effectively harness the potential of AI-driven technologies, such as virtual influencers, in their marketing strategies.

3. Data

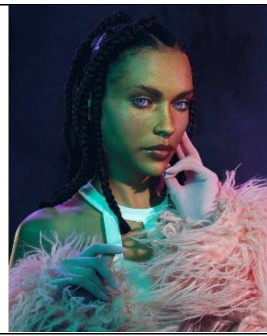
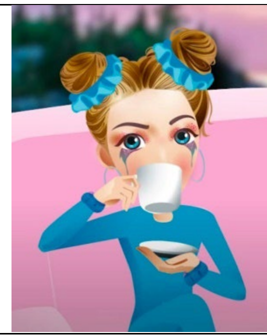
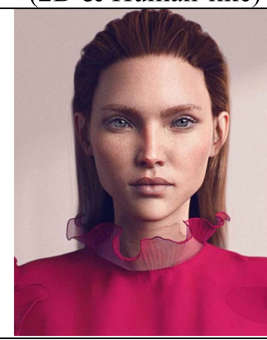
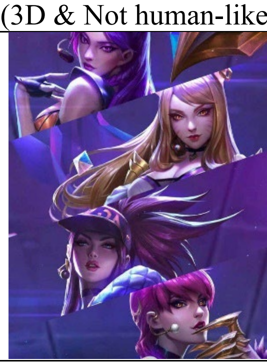
We employ Instagram post data created by virtual influencers[42] and Instagram comment data posted by the Instagram users.

3.1. Identify Virtual Influencers with Sufficient Number of Comments

We acquire a list of virtual influencers through VirtualHumans.com,[43] a third-party influencer marketing company that posts up-to-date news, research, interviews, and biographies about virtual influencers. To examine how Instagram users respond to the virtual influencers' posts, we only include virtual influencers who have been receiving sufficient amount of comments. Among the 168 virtual influencers who are listed by VirtualHumans.com, we include 24 virtual influencers: ai_angelica, alizarexx, annaoop.yt, astrolovesu, bee_nfluencer, birdsoup, bodybyralph, dagny.gram, fnnxrmal, galaxia.gram, guggimon, janky, john.pork, kda_music, koffi.gram, kyraonig, lilmiquela, livinthefuture, noonoouri, realqaiqai, shudu.gram, squeakyandroy, teflonsega, warnymph. These 24 virtual influencers are either human-like (16; 66.7%) or not human-like such as animals or aliens (8; 33.4%). Also, 18 out of 24 virtual influencers are 3D-animated, while the remaining 6 influencers are 2D characters. To avoid any impact from the brands, we eliminated brand-owned virtual influencers such as Totino's. The exhaustive list of virtual influencers is presented in Figure 1.

Figure 1. List of 24 Virtual Influencers

[42] Our Instagram post data and metadata were provided by the Facebook CrowdTangle team.
[43] VirtualHumans: www.virtualhumans.org

			
ai_angelica (3D & Human-like)	alizarexx (3D & Human-like)	Annaoop.yt (2D & Human-like)	astrolovesu (2D & Human-like)
			
bee_nfluencer (3D & Not human-like)	Birdsoup (2D & Not human-like)	bodybyralph (2D & Human-like)	dagny.gram (3D & Human-like)
			
Fnnxrmal (3D & Not human-like)	galaxia.gram (3D & Not human-like)	guggimon (3D & Not human-like)	janky (3D & Not human-like)
			
john.pork (3D & Not human-like)	kda_music (3D & Human-like)	koffi.gram (3D & Human-like)	Kyraonig (3D & Human-like)

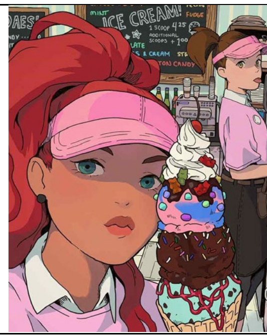
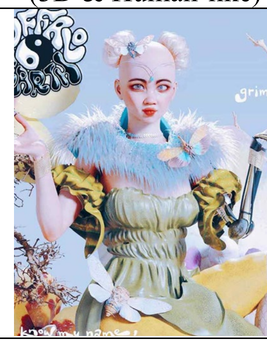
			
lilmiquela (3D & Human-like)	livinthefuture (2D & Human-like)	noonoouri (3D & Human-like)	realqaiqai (3D & Human-like)
			
shudu.gram (3D & Human-like)	Squeakyandroy (3D & Not human-like)	Teflonsega (2D & Human-like)	Warnymph (3D & Human-like)

Table 1 provides the summary statistics of virtual influencers' Instagram popularity, posting behaviors, comments, and engagement metrics during the reference period. We include all the 24 virtual influencers' posts and comments on their Instagram accounts to incorporate how the number of new and revisiting users grow on virtual influencers' Instagram posts. Here, the new users are the users who have never commented before, while the revisiting users are the users who have commented. On average, 24 virtual influencers have accrued 346,260.3 followers most recently, received 26,417.98 likes and 344.683 comments.

Table 1. Summary Statistics of Virtual Influencer Popularity, Posting Behaviors, Comment-related Metrics and Comment Valence

Popularity Metrics	N	Mean	SD	Min.	Max.
Follower count (by influencer)	24	346,260.3	642,964.2	2,587	2,867,204
Like count (by post)	4,762	26,417.98	35,629.08	0	434,079
Comment count (by post)	4,762	344.683	566.484	2	9,505

Posting Behaviors					
Launch year (by influencer)	24	2018.41	1.84	2011	2020
N. of posts (by influencer/year)	4,762	92.14	132.53	1	1,058
Post text length (by post)	4,762	188.496	194.267	0	2,087
Comments					
N. of comments (by post)	4,762	297.047	471.453	2	8,127
N. new users' comments (by post)	4,762	133.832	282.953	0	5,265
N. revisitors' comments (by post)	4,762	102.161	136.263	0	2,220
Comment valence					
Compound valence (by post)	4,762	0.290	0.168	-0.519	0.787
Positive valence (by post)	4,762	0.267	0.113	0.000	0.658
Neutral valence (by post)	4,762	0.665	0.099	0.262	1.000
Negative valence (by post)	4,762	0.068	0.048	0.000	0.548
Notes. The follower count includes missing values as our data does not keep track of the follower count of Instagram accounts with fewer than 50 followers. We calculate the number of posts per year as the influencer's total posts divided by the number of months since joining Instagram.					

3.2. Human Qualities on Virtual Influencers

On their Instagram posts, virtual influencers exhibit their human qualities to make the audience feel like they are interacting with real human beings. We aim to examine the comprehensive impact of exhibiting the human qualities on the consumer engagements on Instagram posts. Specifically, we test three types of human qualities: (1) visual characteristics – image aesthetics and the estimated demographics and emotions of virtual influencers, (2) the estimated human-like pose, and (3) the interaction between virtual influencers and Instagram users.

3.2.1. Visual Characteristics

We calculated the two types of visual characteristics of the virtual influencers, the image aesthetics of influencer content and their estimated demographics (age, gender, race) and emotions. We use the DeepFace (Taigman et al., 2014), a deep-learning-based facial attribute analysis framework from the Facebook research team, to predict their age based on appearance, each virtual influencer's gender, and to predict race (six categories: Asian, Black, Indian, Latino/Hispanic, Middle Eastern, or Non-Hispanic White). Table 2 provides the summary statistics of virtual influencers' visual characteristics.

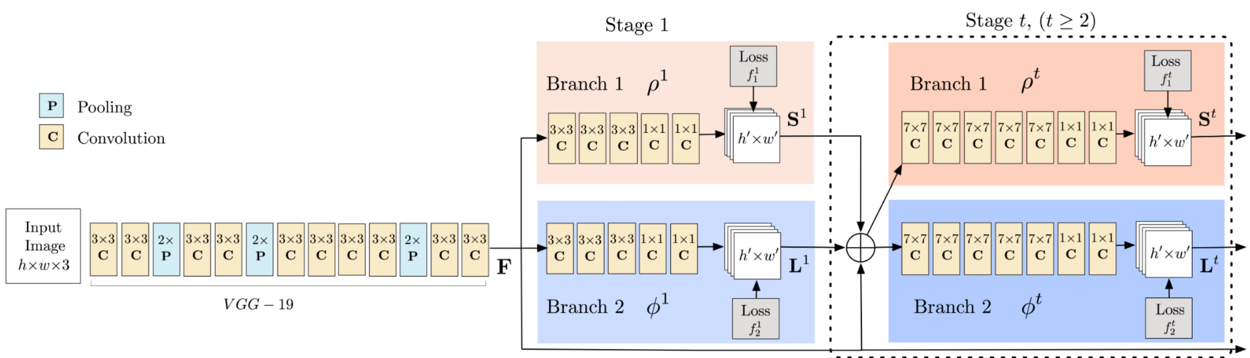
Table 2. Summary Statistics of Virtual Influencer Demographics, Emotions, and Image Aesthetics

Demographics	N	Mean	SD	Min.	Max.
Predicted age (by post)	2,296	28.96	3.227	19.00	48.50
Probability of female (by post; %)	2,296	64.62	40.50	0.00	100.00
Race – Asian (by post; %)	2,296	17.647	22.776	0.00	100.00
Race – Black (by post; %)	2,296	10.819	26.086	0.00	100.00
Race – Indian (by post; %)	2,296	4.469	5.699	0.00	100.00
Race – Latino/Hispanic (by post; %)	2,296	14.717	9.788	0.00	100.00
Race – Middle eastern (by post; %)	2,296	11.028	9.724	0.00	100.00
Race – White (by post; %)	2,296	41.32	29.833	0.00	100.00
Emotions					
Emotion – angry (by post; %)	2,296	4.721	13.558	0.00	100.00
Emotion – disgust (by post; %)	2,296	0.19	2.380	0.00	100.00
Emotion – fear (by post; %)	2,296	8.054	18.511	0.00	100.00
Emotion – happy (by post; %)	2,296	15.697	30.092	0.00	100.00
Emotion – neutral (by post; %)	2,296	50.819	40.695	0.00	100.00
Emotion – sad (by post; %)	2,296	16.536	26.780	0.00	100.00
Emotion – surprise (by post; %)	2,296	3.983	15.932	0.00	100.00
Image aesthetic score (by post)	4,762	5.740	0.657	3.214	7.123

3.2.2. Human-Like Pose Estimation

In addition to the virtual influencers' visual aesthetic variables, their actions and gestures could also affect consumer engagements. More importantly, it is not clear whether and how exhibiting the human-likeness in virtual influencers' behaviors influences consumer engagements. To quantify the degree that virtual influencers' motions and gestures, we employ Realtime Multi-Person Pose Estimation using Part Affinity Fields (PAF; Cao et al., 2017) that leverages two branch multi-stage CNN. Intuitively, Cao et al. (2017) first generates feature maps via 10-layered VGG-19 network, to produce a set of detection confidence maps and part affinity fields. Then in the subsequent stages, the model refines the predictions of the body part by minimizing the loss functions at the two branches. Figure 2 displays the two-branch multi-stage CNN architecture from Cao et al. (2017).

Figure 2. Architecture from the two-branch multi-stage CNN borrowed from Cao et al. (2017)



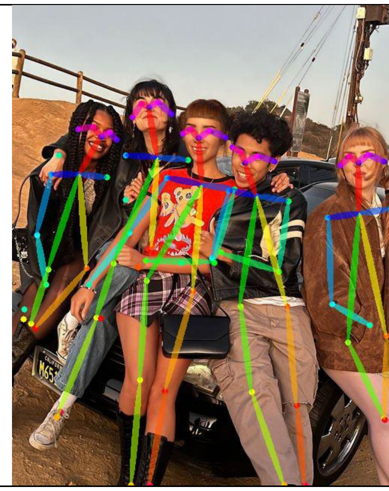
By incorporating Pose Estimation using Part Affinity Fields (PAF; Cao et al., 2017), we detect virtual influencers' body parts that consist of 18 key body points – nose, neck, left & right shoulders, elbows, wrists, hips, knees, ankles, eyes, and ears. Once we detect the eighteen body points, we save the 3D metrics of virtual influencer body parts, and incorporate them to represent the body motions of virtual influencers. Figure 3 illustrates how we detect virtual influencers' body parts, and Figure 4 displays four examples of virtual influencers' body parts that show different motions and gestures detected based on the pose estimation using PAF.

Figure 3. Illustration of Detecting Virtual Influencer (Lil Miquela) Body Parts

Original Image	Detect the right arm and the right knee	Detect 18 body points	Visualize linked body parts

Figure 4. Four Examples of Different Motions and Gestures of Virtual Influencers

A human-like virtual influencer looking in front of the mirror	A human-like virtual influencer sitting on a car
	
Two human-like virtual influencers, one stretching arms and the other holding the wall	Five human-like virtual influencers standing next to each other

By detecting the body parts, we acquire the matrix of the virtual influencer motion data and incorporate them as our independent variables.

3.2.3. Human-Virtual Influencer Interaction

Building the relationship between human users and virtual influencers can be facilitated through their communication on Instagram posts. To understand the impact of the Instagram comment communication, we incorporate the valence and the number of comments written by the virtual influencers as the metrics to reflect the human-virtual influencer interaction. Table 3 provides the summary statistics of virtual influencers' interaction with human users.

Table 3. Summary Statistics of Virtual Influencer Interaction

Virtual Influencer Interaction	N	Mean	SD	Min.	Max.
Number of Influencer Comments	4,762	20.746	32.781	0	287

4. Model

Our research objective is to empirically estimate the effect of virtual influencers exhibiting human qualities on consumer engagement, specifically in terms of user comments. To achieve this, we employ a rigorous analytical approach by estimating two models as outlined in Sections 4.1 and 4.2.

4.1. Changes in the Number of Comments and Comment Sentiment

The extent to which virtual influencers convey human-like qualities in their posts is likely to have varying effects on user engagement, including the number of comments received and consumer sentiment. To examine the impact of human qualities on these dependent variables, we operationalize three distinct dimensions of human qualities, namely (1) visual characteristics encompassing virtual influencers' predicted age, gender, race, emotions, and image aesthetic score, (2) estimated human-like poses, and (3) the frequency of comments written by virtual influencers on their own posts. We posit that virtual influencers may elicit similar responses from users as human influencers do, as documented in previous research (Colombo et al., 2019). We aim to empirically test these hypotheses through rigorous statistical analysis and provide insights into the differential impact of human qualities on virtual influencer-user interactions in Equation (1) and (2).

$$\log(1 + Number\_of\_comments_{ijt}) = \beta_0 + \Gamma' * \overrightarrow{HumanQualities_{ijt}} + I_i + \tau_t + \epsilon_{ijt} \tag{1}$$

$$Comment\_Sentiment_{ijt} = \beta_0 + \Gamma' * \overrightarrow{HumanQualities_{ijt}} + I_i + \tau_t + \epsilon_{ijt} \tag{2}$$

where $Number\_of\_comments_{ijt}$ and $Comment\_Sentiment_{ijt}$ are the number of comments and comment sentiment that influencer *i* gets on post *j* at time *t*.

4.2. Growth of the New and Revisiting Users who Leave Comments

Furthermore, we extend our analysis by investigating the dynamics of user growth among new and revisiting users in response to different human qualities exhibited by virtual influencers. Once again, we consider virtual characteristics, human-like poses, and comment participation as the human quality variables, hypothesizing that the growth of new and revisiting users may be differentially impacted by these human qualities. For instance, we posit that new users may be more likely to leave comments when interacting with a virtual influencer who is new and less familiar, as they seek to establish a connection. On the other hand, revisiting users may be more inclined to leave comments when they feel a sense of attachment to the virtual influencer. To empirically test our hypothesis, we implement two distinct models, denoted as Equation (3) and (4), to examine the relationship between human qualities and the growth of new and revisiting users.

$$\log(1 + Number\_of\_new\_users_{ijt}) = \beta_0 + \Gamma' * \overrightarrow{HumanQualities_{ijt}} + I_i + \tau_t + \epsilon_{ijt} \tag{3}$$

$$\log(1 + Number\_of\_revisiting\_users_{ijt}) = \beta_0 + \Gamma' * \overrightarrow{HumanQualities_{ijt}} + I_i + \tau_t + \epsilon_{ijt} \tag{4}$$

where $Number\_of\_new\_users_{ijt}$ is the number of newly-appearing comment writers on Instagram post j at time t, and $Number\_of\_revisiting\_users_{ijt}$ is the number of revisiting comment writers.

5. Results

Section 5.1 reports the regression results where the dependent variables are the logarithm of the number of comments and the comment sentiment. Section 5.2 presents the regression results where the dependent variables are the logarithm of the number of new and revisiting comment writer users.

5.1. Impact of Human Qualities on Comments

To answer our first question on how human qualities of virtual influencers impact consumer engagements, we interpret the changes in the number of comments and comment sentiment as indicators of consumer engagement. We find that influencers are more likely to receive higher number of comments when their image aesthetics are perceived as more pleasing, they identify as female or Asian, and increase their participation in writing comments on their posts. Conversely, being identified as black is associated with receiving fewer comments. Additionally, we find that influencers can consistently enhance comment sentiment by writing more comments on their own Instagram post. However, if influencers already have a substantial following, it may have a negative impact on comment sentiment. Our findings tell us that virtual influencers have the ability to engage human users when they respond in a manner that emulates human interaction, such as writing more comments. This underscores the significance of human-like responses in fostering user engagement with virtual influencers.

Table 7. ATE Estimations from the DR-DiD Estimator

	Estimates (Std. Err.)			
Variable	**(1) DV: Log of the number of comment**		**(2) DV: Comment Sentiment**	
Image aesthetics	**0.1267*****	**(0.0369)**	-0.0063	(0.0044)
Number of people on Image	0.0230	(0.0166)	0.0004	(0.0031)
log(age)	-0.0154	(0.1679)	0.0004	(0.0191)
Emotion - angry	0.0001	(0.0014)	0.0002	(0.0001)
Emotion - disgust	-0.0032	(0.0060)	-0.0009	(0.0007)
Emotion - fear	0.0015	(0.0010)	-0.00003	(0.0001)
Emotion - happy	-0.0008	(0.0006)	0.00005	(0.0001)
Emotion - sad	0.0003	(0.0007)	-0.00001	(0.0001)
Emotion - surprise	-0.0003	(0.0011)	0.0001	(0.0001)
Gender - female	**0.0018*****	**(0.0006)**	0.0002*	(0.0001)
Race - asian	**0.0026*****	**(0.0008)**	0.0001	(0.0001)
Race - black	**-0.0028****	**(0.0013)**	-0.000002	(0.0002)
Race - indian	-0.0029	(0.0046)	0.0008*	(0.0005)
Race - latino_hispanic	0.0037	(0.0029)	-0.00003	(0.0003)
Race - middle_eastern	0.0014	(0.0027)	-0.0002	(0.0003)
log(1+Followers_at_Posting)	0.0067	(0.0093)	**-0.0025*****	**(0.0010)**
log(1+comment_participation)	**0.1510*****	**(0.0158)**	**0.0136*****	**(0.0021)**
Constant	4.8859***	(0.7990)	0.1128*	(0.0676)

Observation	1,784	1,784
Fixed effects	Influencer and time fixed effects	
R^2	0.7493	0.6282
Notes. All dependent variables are logged. *p<0.1; **p<0.05; ***p<0. 01		

5.2. Impact of Human Qualities on the Growth of New and Revisiting Users on Comment

In addition to examining the impact of human qualities on the number and sentiment of comments, we also investigate their influence on the growth of new and revisiting comment writers. Our findings reveal that virtual influencers who possess a more aesthetic image, have Asian-oriented facial traits, and actively engage in writing Instagram post comments tend to receive higher levels of user engagement in the form of comments. Furthermore, we observe that new users are more likely to engage in leaving comments when there are higher levels of user engagement on the Instagram post image, when the influencer is female, and when the influencer has a smaller following. Conversely, revisiting users are more likely to leave comments when they perceive fear emotion on the influencer's face, when the influencer has a larger following, and when the influencer is not of black ethnicity. These results underscore the significance of virtual influencers actively writing comments to foster user engagement. Interestingly, we also find that user preferences regarding the popularity of virtual influencers differ, with new users showing a preference for smaller-sized influencers, and vice versa.

Table 10. Influencer-Level HTE Results: Demographics

	Estimates (Std. Err.)			
	(1) DV: Log of the number of new users		**(2) DV: Log of the number of revisiting users**	
Image aesthetics	**0.1213*****	**(0.0465)**	**0.1345*****	**(0.0320)**
Number of people on Image	**0.0469****	**(0.0185)**	0.0054	(0.0123)
log(age)	-0.0357	(0.2109)	-0.0926	(0.1598)
Emotion - angry	0.0005	(0.0016)	-0.0005	(0.0010)
Emotion - disgust	-0.0038	(0.0072)	-0.0005	(0.0039)

Emotion - fear	0.0005	(0.0013)	**0.0017****	**(0.0009)**
Emotion - happy	-0.0008	(0.0007)	-0.0007	(0.0005)
Emotion - sad	0.0002	(0.0009)	0.0004	(0.0006)
Emotion - surprise	0.0001	(0.0013)	-0.0015	(0.0012)
Gender - female	**0.0023*****	**(0.0007)**	0.0008	(0.0005)
Race - asian	**0.0038*****	**(0.0010)**	**0.0023*****	**(0.0008)**
Race - black	-0.0027*	(0.0015)	**-0.0030****	**(0.0012)**
Race - indian	-0.0046	(0.0052)	-0.0039	(0.0033)
Race - latino_hispanic	0.0054	(0.0035)	0.0025	(0.0024)
Race - middle_eastern	0.0011	(0.0035)	0.0009	(0.0022)
log(1+Followers_at_Posting)	**-0.0294*****	**(0.0097)**	**0.0349*****	**(0.0084)**
log(1+comment_participation)	**0.0788*****	**(0.0205)**	**0.0674*****	**(0.0143)**
Constant	4.0889***	(0.9518)	-0.1747	(1.5344)
Observation	1,784		1,784	
Fixed effects	Influencer and time fixed effects			
R^2	0.7123		0.7639	
Notes. All dependent variables are logged. *p<0.1; **p<0.05; ***p<0. 01				

6. Conclusion

In this paper, we investigate the impact of human qualities exhibited by virtual influencers on various metrics of consumer engagements. Specifically, we incorporate user comments, comment sentiment, and the growth of new and revisiting comment writers. Our results highlight that virtual influencers are more likely to receive higher numbers of comments when they are perceived to have more pleasing image aesthetics, identify as female or Asian, and actively participate in writing comments on their posts. Conversely, being identified as black is associated with receiving fewer comments. Furthermore, we find that virtual influencers can consistently enhance comment sentiment by actively engaging in writing comments on their own Instagram posts, although this effect may be moderated by the size of their following. Our findings underscore the importance of virtual influencers responding in a manner that emulates human interaction, such as actively writing comments, to foster user engagement.

In addition, we show that virtual influencers with a more aesthetic image, Asian-oriented facial traits, and writing more comments on their posts tend to receive higher levels of user engagement in the form of comments. Furthermore, we observe that user preferences regarding virtual influencers' popularity differ, with new users showing a preference for smaller-sized influencers, and revisiting users showing a preference for influencers with larger followings. We illustrate the nuanced nature of consumer engagement with virtual influencers and provide valuable insights for marketers and practitioners seeking to optimize virtual influencer strategies.

Our research contributes to the growing body of literature on virtual influencers and consumer engagement by providing empirical evidence of the impact of human qualities on user comments, comment sentiment, and the growth of comment writers. We emphasize the significance of virtual influencers actively engaging in human-like responses, such as writing comments, to foster user engagement and we reveal the nuanced preferences of different user segments. Further research can explore other contextual factors that influence consumer engagements with virtual influencers, and how these insights can be leveraged to optimize virtual influencer marketing strategies.

References

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7291-7299).

Chen, Y., Wu, S., & Guo, X. (2019). Trust in social robots: A multidisciplinary review. International Journal of Human-Computer Interaction, 35(10), 901-923.

Choi, Y., Lee, J., & Sung, Y. (2021). Authenticity of AI-generated content in influencer marketing: The roles of perceived deception and consumer characteristics. Journal of Advertising, 50(2), 196-211.

Gupta, P., Das, S., Jain, S., & Garg, R. (2020). Virtual influencers and consumer behavior: An exploratory study on consumers' perception and attitudes. Journal of Retailing and Consumer Services, 53, 101971.

Kim, S. K., Cho, S. E., & Lee, H. J. (2020). The effects of explainable artificial intelligence (XAI) on consumer trust and adoption intentions. Computers in Human Behavior, 110, 106380.

Lu, T., Shen, C., Huang, L., & Chen, T. (2019). Virtual influencers on social media: Analyzing follower comments from parasocial interaction perspective. In Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS), 4383-4392.

Marwick, A. E., & boyd, d. (2019). The drama! Teen conflict, gossip, and bullying in networked publics. Human Behavior and Emerging Technologies, 1(1), 6-19.

Moustakas, E., Lamba, N., Mahmoud, D., & Ranganathan, C. (2020, June). Blurring lines between fiction and reality: Perspectives of experts on marketing effectiveness of virtual influencers. In 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security) (pp. 1-6). IEEE.

Stein, J. P., Linda Breves, P., & Anders, N. (2022). Parasocial interactions with real and virtual influencers: The role of perceived similarity and human-likeness. New Media & Society, 14614448221102900.