

Unravelling Consumer Engagement in Influencer Marketing: An Exploratory UGC Analysis

Yi-Ching Tseng^{1,†}
Pei-Yi Chen^{2,†}
Jing Ren^{3,*}
Yu-Chen Hung⁴
Wenting Liu⁵
Jorng-Tzong Horng^{6,*}
Li-Ching Wu⁷

*Corresponding author

¹ National Central University, Taiwan, 109502515@cc.ncu.edu.tw

² National Central University, Taiwan, 109502504@cc.ncu.edu.tw

³ Lecturer, Singapore University of Social Sciences, Singapore, jingren@suss.edu.sg

⁴ Senior Lecturer, Singapore University of Social Sciences, Singapore, ychung@suss.edu.sg

⁵ Senior Lecturer, Singapore University of Social Sciences, Singapore, wentingliu@suss.edu.sg

⁶ Professor, National Central University, Taiwan, horng@db.csie.ncu.edu.tw

⁷ Associate Professor, National Central University, Taiwan, nculcwu@gmail.com

† Contributed equally

ABSTRACT

The sharp increase in the influencer market on social media generates a huge amount of content, including text, audio, images, and video. However, how these contents are used to leverage consumer engagement is still unclear. In this study, we explored the potential usage of natural language processing and various computer vision models in understanding influencer UGC and marketing analysis. Through studying the influencers' textual and visual posts on Instagram, our preliminary findings show that different formats of UGC may impact consumer engagement differently. Our work offers new knowledge on exploring suitable analysis tools for studying the influencer market, and the learned insights on content generation could shed light on enhancing marketing effectiveness in the influencer marketing industry.

Keywords: computer vision, consumer engagement, influencer marketing, LIWC, and UGC.

INTRODUCTION

With a current market value of 10.39 billion USD, an annual 33.4% increase from 2019, influencer marketing on social media is a very young but fast-expanding industry. Social media influencers are individuals who have gained popularity because of their presence and content on social media, such as bloggers, YouTubers, TikTok Live users, and Instagrammers (Aw and Chuah, 2021), and they are cashing in on their popularity to help brands promote goods and services. Today, social media influencers are not just mega-influencers with follower counts in the millions; they could also be micro-influencers with dedicated followers and niche knowledge. They can include both conventional celebrities who first became famous outside of social media as well as those who became famous fully online (Campbell and Farrell, 2020).

In this emerging business model, influencers partner with brands to craft interesting and impactful content that is relevant to their followers. Through the partnership, brands aim to increase engagement via influencers, specifically in light of their interactivity (Farivar *et al.*, 2022; Jun and Yi, 2020). The beauty of this system is that influencers, rather than the brand itself, can be the generators of marketing content, which is usually called User-Generated Content (UGC). A common format of UGC is visual content (e.g., image, video) plus textual content.

To promote a product using visual content such as images, influencers may use objects to express themselves innovatively. Objects often carry symbolic meanings, acting as a metaphor that evokes emotions. Their relationship with the human subjects and other objects in an image can convey meaning and feelings among the readers. For textual content, influencers may have their own writing styles, such as adding more tags, using a personal tone, using long sentences, etc. Despite the proliferation of these practices, it remains unclear how and to what extent the generated content could create social engagement with its target audience, especially by linking to products in different categories. To our knowledge, the bulk of the research on the impact of visual and textual popularity prediction on consumer engagement is still insufficient.

In this study, we seek to leverage a hybrid exploratory analysis using natural language processing and computer vision models to understand influencer-generated visual and textual content. By exploring the impact of image features and textual features on engagement rates, we aim to answer three research questions:

- What kind of image or text could drive a great number of “like” engagements with an influencer’s post?
- What kind of image or text could drive a great number of “comment” engagements with an influencer’s post?
- Compared to human annotation, could a machine outperform in predicting consumer engagement?

By studying the UGC content posted through collaborations between brands and influencers on Instagram for reputation management and special promotions, our work plans to offer new knowledge on exploring suitable analysis tools for studying the influencer market. It also paves the way for enhancing the effectiveness of influencer marketing and providing suggestions for sustainable career development for influencers.

LITERATURE REVIEW

Influencer Marketing

Brands have gradually abandoned traditional advertising and increasingly use social media influencers to promote their products or services (Janssen et al., 2022). The market value of influencer marketing platforms was estimated to be with an expected expansion at an annual growth rate of 33.4% until 2030 across campaign management, search and discovery, analytics and reporting, and influencer management on popular social media sites such as Instagram, Meta (Facebook), Pinterest, and YouTube (Grand Viewer Research, 2021). When brands partner with influencers, they compensate the influencers for their services, including content creation, posting on their channels, attending events, providing feedback or reviews, or even selling products. In return, influencers may receive a flat fee, commission, free products or services, or sponsorship for content generation in exchange for their promotion on social media. In the partnership, brands use influencer marketing as an effective tool for digital marketing communication, where interactivity and engagement are the core to delivering value (Farivar et al., 2022).

To generate engagement (e.g., like, comment, and share), influencers rely on creative content and interactive features. The content needs to be visually appealing to grab the attention of their followers, such as interesting photos, short videos, and captive captions. The visual inputs spark conversation, and influencers also use hashtags to co-create with other influencers for brands. Using trending hashtags, influencers can reach a wider audience through timely online conversations with other influencers and brand followers. Trendy and attention-grabbing visuals are the key characteristics of posts by Instagram influencers. Visual inputs of interesting and unexpected images serve as a foundation to simulate engagement with followers.

Different social media platforms have varying media formats, purposes, and audiences. In a comparative study of popular social media platforms, Haenlein et al. (2020) found that Facebook tends to have an older audience with lower engagement rates. Twitter supports text-based posts only and is often utilized as a news source. Additionally, both YouTube and Instagram are primarily used for entertainment purposes, as pursuing entertainment value can facilitate the expansion of influence. Given the aforementioned advantages—namely, a younger user demographic, diverse UGC formats, and the potential for easy influence extension—this study opts for influencer marketing posts on Instagram. The aim is to investigate which types of UGC are favored by consumers, with the subsequent goal of enhancing advertising engagement.

UGC Analysis in Consumer Engagement

Textual content analysis

Text analysis techniques have been adopted in various applications, including social media engagement (Hung and Guan, 2020). Language and communication styles, including textual features such as post length, structure, personal tones, and word choice, may serve as persuasive cues for consumer engagement on social media platforms. A study by Fan et al. (2023) investigated the top 10 influencers on Sina Weibo, which is the leading microblogging social media platform in China. They suggested that including promotional messages in the post has the potential to boost consumer engagement; however, they did not learn the textual content in depth. Syrdal et al. (2023) used Linguistic Inquiry and Word Count (LIWC) to extract linguistic-based features from the textual post of Instagram, and they found that using words with high authenticity and positive emotions in the post can increase consumer engagement. About the length of the post, these two studies hold different conclusions. Fan et al. argued that longer text provides richer information or incorporates non-commercial messages into a story, leaving a deeper impression on the audience. While Syrdal et al. suggested that longer and more logical posts may discourage audiences from understanding the details.

Although consumer engagement with different cultural backgrounds may give different responses to the length of the post, the textual content has been recognized as an important item for formulating marketing content and strategies. However, how the brand content aligns with the influencer's style, leading to a higher consumer engagement rate, is still worth finer-grained investigations.

Visual content analysis

Exploration of the impact of visual UGC (e.g., image, video) is attracting more and more attention with the sharp increase of influencer marketing in the marketing field. However, the grouping amount and diversity of UGC make manual coding annotation hard to handle. The interest in automatic detection, coding, or labeling using computer vision (CV) techniques has

become a hot topic in recent years (Kwon et al., 2022). In principle, CV models, especially deep neural networks, could provide strong support to brand-related visual UGC analysis for marketing applications because CV can automatically analyze the visual content, for example, extract color and texture, detect and classify objects, and classify scenes and themes. Unfortunately, its application is still in the early stages, on either a research or practical level. A possible obstacle is the lack of pre-trained CV models for brand-related visual content. The existing research explores the possibility and performance of using CV models to gain marketing concepts, which can be summarized into three branches:

- Based on low-level features: Kwon et al. (2022) revealed that various patterns of color features (e.g., cool or warm, darkness, saturation, and colorfulness) have been used in B2B and B2C contexts to gain positive eWOM on social media platforms such as Instagram.
- Based on commercial software: A bulk of existing work adopted commercial software, such as the Google Cloud Vision API and Clarifai API. For example, the Google Cloud Vision API has been used to label tourist-generated images and food images to understand the impact of visual content on consumer engagement (Philp et al., 2022). Nanne et al. (2020) verified the ability of Clarifai to analyze images on Instagram. The disadvantage of using commercial software is that developers have little control over the method's design.
- Based on pre-trained CV models: The tested CV model is YOLOv2, but it did not show a promising outcome (Nanne et al., 2020). Unfortunately, other famous CV models like OpenCV, Show&Tell, and generative AI-based ClipCap have not been tested (Vinyals et al., 2015; Mokady et al., 2021).

Hybrid analysis

A small group of researchers has tried to study the ability of combining textual and visual content to predict consumer engagement in influencer marketing. For example, Jaakonmäki et al. (2017) integrated basic textual features (e.g., words and emojis) and visual features (i.e., detected objects) using the Clarifai API. However, they did not learn the text in depth and ignored illustrating the impact of the learned visual features. Nanne et al. (2020) tried one more step to involve topic modeling to group the visual words learned by YOLOv2 and commercial tools; however, they did not study the textual content.

DATA CONSTRUCTION

The goal of this study is to explore the role of user-generated textual and visual content in predicting consumer engagement on social media. We collected UGC posts published by Instagram influencers from a collaboration marketing company. All the influencers' posts are in collaboration with brands. There are a total of 169 UGC posts involving six brand categories: *Apparel, Cosmetics, E-commerce, Food, Hotels, and Watches*. The involved brands are Calvin Klein, Nike, Prada, Shiseido, SKII, Lotte, Burger King, McDonald's, Ritz Hotel, Daniel Wellington, Rolex, etc.

Each UGC post contains images and text. The image content could include the brand product, and the textual content tends to be emotional, often reflecting influencers sharing their personal experiences using the products or expressing gratitude to the brand for the business collaboration. The datasets also encompass basic information about the influencers and their engagement metrics (such as likes and comments), enabling further content analysis and influence assessment. In this section, we will provide a detailed explanation of the dataset's construction and data description.

UGC Analysis and Measures

Textual variables

Two elements of the textual features are extracted based on the literature review: basic information and linguistic features. Basic information includes the length of the post (*postLength*), whether the product brand appeared (*brandNameAppeared*), and whether the product category appeared (*productCategoryAppeared*). For linguistic features, LIWC is adopted, which is a tool for analyzing linguistic features or emotional tendencies in text based on linguistics and psychology. It aims to provide quantitative and qualitative analysis of textual content (Pennebaker et al., 2001). Existing studies reflected that language and communication styles could be persuasive factors for consumer engagement; therefore, we extracted two main dimensions of linguistic features using LIWC:

- Personal pronouns: The analysis of pronoun categories of influencers' posts can indicate which pronoun categories are used more frequently or less frequently in the text. For example, extensive use of the first-person singular pronouns may indicate an emphasis on personal perspectives, experiences, or emotions. While the use of second-person pronouns may suggest interaction with the reader or providing instructions. In total, six dimensions are learned, including a general pronoun and five sub-categories to indicate the writing style in using different pronouns (as shown in Table 2).
- Psychological processes: Five psychological process categories, including *affective, social, cognitive, perceptual, and physical*, are extracted. Psychological process categories can examine which categories dominate the language elements of the post. For example, the widespread use of *affective* process categories may indicate an emphasis on emotional expressions such as preferences, concerns, or excitement. While *social* process categories may imply a focus on social interaction or interpersonal relationships. The *perceptual* processes, on the other hand, express human senses, including auditory, visual, and tactile senses. The *physical* processes encompass not only bodily structures but also physiological states and sexual orientation.

Visual variables

Four open pre-trained CV models are used to test the suitability of supporting marketing analysis compared to using manual annotations given by humans. In total, three undergraduate students from the Computer Science discipline were hired to label the images, and the four open pre-trained models are YOLOv5, OpenCV, ClipCap, and Show & Tell. Various evaluation metrics such as BLEU, ROUGE, and METEOR (Vedantam et al., 2015) were adopted to evaluate their performances compared to manual annotations. Considering the potential impact of the influencer's image, we try to compare the detection of key items in an image including humans and objects from 5 dimensions: *#people*, gender of the people (i.e., *woman*, *man*), whether human *face* and other *objects* appeared in the image (the detailed description of the variables is shown in Table 2).

Two labeled samples are given in Figure 1. YOLO and OpenCV are object detection models. They can identify the objects, including persons and items, and also indicate the position of detected objects using different colors of boxes (as shown in Figure 1). While ClipCap and Show & Tell are image descriptions or captioning models with a single sentence. The overall performances can be summarized as follows:

- YOLOv5: It excels at accurately detecting and localizing people but struggles with images containing multiple objects.
- OpenCV: The bounding of the contours of objects is almost completely incorrect, e.g., bounding an area that does not contain any objects, or the accuracy of detecting the precise location of a complete object is quite low, e.g., bounding only one leg of a person. Furthermore, due to the lack of specific classifiers (e.g., a face classifier), OpenCV cannot label the detected objects or generate image descriptions. This misalignment with our research objective.
- ClipCap: As an image captioning model leveraging CLIP (Contrastive Language–Image Pre-training) based on GPT2, it excels at describing relationships between people in images and accurately describing the context of the entire image (e.g., describing the new products on a McDonald's menu), which provides a higher-level understanding of the image content. However, it sometimes cannot determine detailed information about the people and objects.
- Show & Tell: It is capable of describing an image with a single sentence, following the typical sentence structure used by humans to describe images (e.g., "A woman is standing in front of the building"). However, there may be some differences in recognizing the number of people. Nevertheless, the generated descriptions are more aligned with the initially manually annotated sentence structure, so they obtain higher scores in the evaluation metrics.







Input	YOLOv5	OpenCV	ClipCap	Show & Tell
			Person is a fashion actor	A man in a white shirt and jeans is standing in front of a brick wall
			Person and actor are photographed for the wrap	A group of people are standing outside of a building

Figure 1: Visual content labeling using computer vision models.

Although each model has its strengths and weaknesses, the evaluation scores based on the manually annotated sentence structure explained that Show & Tell had the highest scores for the whole dataset (with an average BLEU score of 0.5363, an average METEOR score of 0.2893, and an average ROUGE score of 0.5775), which indicates its suitability for understanding the visual UGC generated by influencers. Therefore, the learned visual variables will be used in the following analysis of consumer engagement.

Engagement Rate and Data Description

Consumer engagement on Instagram is commonly measured using the number of likes (*#Like*) and the number of comments (*#Comment*). Table 1 shows the descriptive statistics. Considering the huge variance in the number of influencers' followers (*#follower*), the engagement rates for influencer content are calculated for the prediction analysis (Syrdal et al., 2023):

- $L\text{-ratio} = \#Like / \#follower$
- $C\text{-ratio} = \#Comment / \#follower$
- $LC\text{-ratio} = (\#Like + \#Comment) / \#follower$

Table 1: Descriptive statistics on engagement & engagement rate.

	MIN	MAX	MEAN	SD
<i>#Like</i>	0	10,002,079	912,832	2,357,808
<i>#Comment</i>	1	2,520,969	24,809	195,618
<i>#follower</i>	6,620	83,800,000	10,889,190	22,048,114
<i>L-Ratio</i>	0	0.532	0.059	0.076
<i>C-Ratio</i>	4.86-E5	0.536	0.061	0.081
<i>influencerType</i>	42 virtual (24.9%), 127 human (75.1%)			
Notes. Obs.: 169 UGC posts including image and text, covering 6 categories: Apparel (92), Cosmetics (17), E-commerce (2), Food (23), Hotels (16), and Watches (19).				

Table 2 lists all the 21-D variables that will be used in the following engagement analysis, which include 14-D textual variables, 5-D visual variables, and 2-D control variables. *#brandFollower* was included to control the possible impact coming from the popularity of the product or brand, and *influencerType* was also involved to moderate the possible impact of different types of influencers.

Table 2: Variable description.

NOTATION	CONTEXT INFO
Text Variables	
<i>brandNameAppeared</i>	Whether contains the product brand name, Yes(1) or No (0)
<i>productCategoryAppeared</i>	Whether contains the product category, Yes(1) or No(0)
<i>lengthOfThePost</i>	The length of the post
<i>personalPronouns</i>	The personal pronouns used in the post writing
<i>i</i>	First-person singular, such as I, me, and mine
<i>we</i>	First-person plural, such as we, us, and our
<i>you</i>	Second-person, such as you and your
<i>shehe</i>	Third-person singular, such as she, her, he, and his
<i>they</i>	Third-person plural, such as they, their, and them
<i>affective</i>	Refers to vocabulary related to emotions and tone of speech
<i>social</i>	Related to daily interpersonal interactions, including social behavior, communication, relationships, and various social references
<i>cognitive</i>	Represents the category of vocabulary related to mental processes, such as understanding, memory, and problem-solving
<i>perceptual</i>	Represents a category that includes vocabulary related to actions, space, and sensory perception
<i>physical</i>	Represents vocabulary related to physiology and bodily sensations
Image Variables	
<i>#people</i>	The number of people in the image
<i>woman</i>	Whether woman is in the image, Yes(1) or No(0)
<i>man</i>	Whether man is in the image, Yes(1) or No(0)
<i>face</i>	Whether the human face is shown in the image, Yes(1) or No(0)
<i>objects</i>	Whether the sponsor's product appeared in the image, Yes(1) or No(0)
Control Variables	
<i>brandFollower</i>	The number of Instagram followers of the brand
<i>influencerType</i>	The type of the influencer, virtual (1) or human (0)
Note. The linguistic values for <i>personalPronouns</i> , <i>affective</i> , <i>social</i> , <i>cognitive</i> , <i>perceptual</i> , and <i>physical</i> are generated using LIWC. A larger value indicates a higher usage frequency or more emphasis on the corresponding dimension.	

RESULTS

Instagram users are exposed to both textual and visual UGC when they open an influencer's post. From the UGC perspective, image content may be easier to attract users' attention because of its position (above the text or on the left of the page) and size

(it usually occupies more space compared to text content). From the users' perspective, it is common knowledge that "Like" is an easier interactive activity compared to giving a "Comment". Therefore, to answer the proposed research questions, we conducted regression models on the three engagement rates, respectively. For each one, we compared the performances of human annotation and a computer vision model (i.e., Show & Tell) in engagement prediction.

LC-ratio

Table 3 gives the regression results. In general, human annotation and computer vision learned similar insights from the impact of textual and visual content on *LC-ratio* engagement, although humans perform slightly better than computer vision in explaining the data (i.e., R^2 value: M1 = 0.330 vs. = 0.301; M1a = 0.341 vs. M2a = 0.312).

In terms of basic textual variables, adding the brand name to the textual content can significantly boost consumer engagement, while adding the product category is not that important. A long post could negatively affect consumer engagement significantly, which is similar to the findings in Syrdal et al.'s study (2023). For linguistic features, we conducted two models by considering the general *personalPronouns* level and the detailed pronoun level, respectively. Both show a significantly negative impact of *personal pronouns* on engagement. Interesting findings are located in *cognitive* and *perceptual* processes. *Cognitive* processes give rise to implicit aspects of thinking, such as causation, insight, and inclusiveness or exclusiveness. Under the influencer UGC context, the significantly positive impact of *cognitive* (p -value < 0.05) reflects that including an explanation of why the product is attractive, e.g., advantages, value, etc., could engage consumers' activity. The *perceptual* processes, on the other hand, express human senses, including auditory, visual, and tactile senses. It shows a significantly negative impact on consumer engagement for all four models (p -value < 0.05), which means mentioning a lot of personal perceptual feelings would be detrimental to engagement.

The finding on *perceptual* factors in the textual content somehow matched the impact of the visual content. The regression results on visual variables show that none of them is significant.

About the two control variables, *#brandFollower* shows different impacts by using human annotation and computer vision. The higher number of brand followers can boost consumer engagement significantly in M1 and M1a (p -value < 0.01) while performing negatively in M2 and M2a, although the impact is not significant.

Table 3: Modeling results of LC-ratio for human annotation and Show&Tell

VARIABLES	TEXT + HUMAN ANNOTATION		TEXT + SHOW&TELL	
	M1	M1a	M2	M2a
<i>Constant</i>	1.217	0.342	2.505	2.365
Textual Variables				
<i>brandNameAppeared</i>	0.238*	0.199 ^o	0.293*	0.250*
<i>productCategoryAppeared</i>	-0.098	-0.051	-0.141 ^o	-0.089
<i>postLength</i>	-0.255**	-0.230*	-0.215*	-0.203*
<i>personalPronouns</i>	-0.288*		-0.326**	
<i>i</i>		-0.206*		-0.236**
<i>we</i>		-0.133		-0.146
<i>you</i>		-0.060		-0.103
<i>shehe</i>		-0.106		-0.124 ^o
<i>they</i>		-0.088		-0.102
<i>affective</i>	0.018	0.028	-0.081	-0.057
<i>Social</i>	0.051	-0.061	0.085	-0.025
<i>cognitive</i>	0.185*	0.202*	0.179*	0.194*
<i>perceptual</i>	-0.250**	-0.212*	-0.206*	-0.169*
<i>physical</i>	-0.057	-0.072	-0.030	-0.048
Visual Variables				
<i>#people</i>	-0.088	-0.071	0.035	0.077
<i>woman</i>	0.266	0.256	0.016	0.021
<i>man</i>	0.269	0.257	-0.005	-0.024
<i>face</i>	-0.131	-0.128		
<i>objects</i>	-0.003	-0.008	-0.086	-0.098
Control Variables				
<i>#brandFollower</i>	0.218**	0.239**	-0.041	-0.017
<i>influencerType</i>	-0.061	-0.016	-0.128	-0.121
<i>R² value</i>	0.330	0.341	0.301	0.312
Notes. ^o $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$				

L-ratio and C-ratio

We further conducted four models to understand *L-ratio* and *C-ratio* using text + human annotations and text + Show&Tell separately. In general, computer vision beats humans in predicting *L-ratio* (e.g., R^2 value: M5 = 0.286 vs. M3 = 0.126), while humans outperform in understanding *C-ratio* compared to computer vision (e.g., R^2 value: M4 = 0.215 vs. M6 = 0.194).

Studying the modeling results of using text + human annotation, visual features present a significant impact on boosting “Like” engagement, especially when people appear in the image (e.g., the influencer, with a p -value <0.05 for males and a p -value < 0.01 for females). However, showing the whole face is not a positive factor for the used data. On the other hand, for *C-ratio*, both textual and visual variables contribute to the engagement to some extent. When a user leaves comments, he or she may take time to read through the text for more information beyond a mere glance at the image. We can see that *brandNameAppeared* and *cognitive* processes are significantly positive, which is the same as their impact on the *LC-ratio*. However, the expression of strong emotion could bring a negative impact because the coefficient of *affective* processes is negative (p -value < 0.01).

Different from using text + human annotation, the findings of the four models using text + Show&Tell are mainly contributed by textual factors. For example, involving more *cognitive* processes and less personal perceptual feeling could boost the *L-ratio* and *C-ratio*. The silence on the visual side could be caused by poor image annotation of computer vision models compared to human beings.

Table 4: Modeling results of L-ratio and C-ratio for human annotation and Show&Tell

VARIABLES	TEXT + HUMAN ANNOTATION				TEXT + SHOW&TELL			
	L-RATIO		C-RATIO		L-RATIO		C-RATIO	
	M3	M3a	M4	M4a	M5	M5a	M6	M6a
<i>Constant</i>	-9656.87	-16460.40	0.071	0.184	2.488	2.402	0.094	0.048
Textual Variables								
<i>brandNameAppeared</i>	-0.082	-0.161	0.213*	0.239*	0.240*	0.194°	0.297**	0.293**
<i>productCategoryAppeared</i>	0.091	0.127	0.070	0.075	-0.145	-0.094	-0.020	0.003
<i>postLength</i>	-0.005	0.059	-0.095	-0.134	-0.192*	-0.172°	-0.143	-0.174°
<i>personalPronouns</i>	0.008		-0.150		-0.318**		-0.140	
<i>i</i>		-0.075		-0.054		-0.235**		-0.084
<i>we</i>		-0.015		-0.108		-0.125		-0.129
<i>you</i>		0.224		-0.158		-0.073		-0.156
<i>shehe</i>		0.029		-0.087		-0.111		-0.086
<i>they</i>		0.006		-0.065		-0.092		-0.065
<i>affective</i>	-0.173°	-0.187°	-0.256**	-0.247**	-0.080	-0.054	-0.050	-0.054
<i>social</i>	-0.074	-0.241	0.045	0.107	0.079	-0.059	0.049	0.127
<i>cognitive</i>	-0.084	-0.093	0.301**	0.326***	0.076	0.085	0.468***	0.493***
<i>perceptual</i>	0.131	0.174°	-0.005	-0.009	-0.195*	-0.156°	-0.100	-0.091
<i>physical</i>	0.072	0.052	0.035	0.044	-0.046	-0.065	0.039	0.042
Visual Variables								
<i>#people</i>	-0.199	-0.189	-0.173	-0.167	0.046	0.086	-0.038	-0.019
<i>woman</i>	0.577**	0.615**	0.350	0.325	-0.013	-0.009	0.126	0.133
<i>man</i>	0.441*	0.441*	0.153	0.154	-0.053	-0.074	0.191	0.197
<i>face</i>	-0.157°	-0.140	0.005	-0.002				
<i>objects</i>	-0.015	-0.020	-0.223**	-0.227**	-0.078	-0.095	-0.048	-0.037
Control Variables								
<i>#brandFollower</i>	-0.069°	-0.045	-0.183*	-0.193*	-0.042	-0.016	-0.009	-0.010
<i>influencerType</i>	0.089	0.152	-0.011	-0.029	-0.136°	-0.131	-0.026	-0.016
<i>R² value</i>	0.126	0.150	0.215	0.220	0.286	0.295	0.194	0.202
Notes. ° $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$								

CONCLUSION

In this study, we explored NLP and CV models for understanding influencer UGC posts and further evaluated the impact of UGC on consumer engagement. The findings provide suggestions for more effective influencer marketing strategies to boost brand awareness. Notably, although visual plus textual content is a common combo to promote products or services, the visual content is effective for generating "Likes," while a combination of visual and relevant textual content can increase "Comments." This insight can guide influencers and brands in understanding their followers and adapting UGC creation. While automatic annotation didn't surpass human labeling, it performed similarly. With the growth of UGC and virtual influencers, this study sheds light on artificial intelligence techniques to aid influencers and sponsors in enhancing brand recognition.

Our work is still in the early stages. We plan to implement a completed study further in the next step. First, our dataset is limited, which may cause conclusion bias. We will collect more data and apply a finer-grained investigation to different categories and

brands of products. Second, consider the impact of the interaction of textual and visual features, for example, the interaction impact of the cognitive process featured in the product category or the writing pronouns and gender of the influencer. Third, improve the performance of CV models to gain a better understanding of visual content. Influencers can tell a story and transmit social significance by including various objects in a shot and arranging them in a specific way. An image of a person surrounded by books, for instance, can suggest that someone enjoys reading or is well-informed. However, the hidden higher-level meanings of the image, e.g., the symbolic meaning of the image and contextual relationships between people, objects, and backgrounds, are all lacking in investigation. We will try to explore the abilities of other CV models, including but not limited to generative AI.

REFERENCES

- Aw, E. C. X., & Chuah, S. H. W. (2021). "Stop the unattainable ideal for an ordinary me!" fostering parasocial relationships with social media influencers: The role of self-discrepancy. *Journal of Business Research*, 132, 146-157. <http://dx.doi.org/10.1016/j.jbusres.2021.04.025>
- Campbell, C., & Farrell, J. R. (2020). More than meets the eye: The functional components underlying influencer marketing. *Business horizons*, 63(4), 469-479. <http://dx.doi.org/10.1016/j.bushor.2020.03.003>
- Fan, F., Chan, K., Wang, Y., Li, Y., & Prieler, M. (2023). How influencers' social media posts have an influence on audience engagement among young consumers. *Young Consumers*, 24(4), 427-444.
- Farivar, S., Wang, F., & Turel, O. (2022). Followers' problematic engagement with influencers on social media: An attachment theory perspective. *Computers in Human Behavior*, 133, 107288. <http://dx.doi.org/10.1016/j.chb.2022.107288>
- Grand Viewer Research (2021). Retrieved from <https://www.grandviewresearch.com/industry-analysis/influencer-marketing-platform-market> (Last accessed June 12th, 2023).
- Haenlein, M., Anadol, E., Farnsworth, T., Hugo, H., Hunichen, J., & Welte, D. (2020). Navigating the New Era of Influencer Marketing: How to be Successful on Instagram, TikTok, & Co. *California management review*, 63(1), 5-25. <http://dx.doi.org/10.1177/0008125620958166>
- Hung, Y. C., & Guan, C. (2020). Winning box office with the right movie synopsis. *European Journal of Marketing*, 54(3), 594-614. <http://dx.doi.org/10.1108/EJM-01-2019-0096>
- Jaakonmäki, R., Müller, O., & Vom Brocke, J. (2017, January). The impact of content, context, and creator on user engagement in social media marketing. In *Proceedings of the Annual Hawaii International Conference on System Sciences* (Vol. 50, pp. 1152-1160). IEEE Computer Society Press. <http://dx.doi.org/10.24251/HICSS.2017.136>
- Janssen, L., Schouten, A. P., & Croes, E. A. (2022). Influencer advertising on Instagram: product-influencer fit and number of followers affect advertising outcomes and influencer evaluations via credibility and identification. *International journal of advertising*, 41(1), 101-127. <http://dx.doi.org/10.1080/02650487.2021.1994205>
- Jun, S., & Yi, J. (2020). What makes followers loyal? The role of influencer interactivity in building influencer brand equity. *Journal of Product & Brand Management*, 29(6), 803-814. <http://dx.doi.org/10.1108/JPBM-02-2019-2280>
- Kwon, J., Chan, K. W., Gu, W., & Septianto, F. (2022). The role of cool versus warm colors in B2B versus B2C firm-generated content for boosting positive eWOM. *Industrial Marketing Management*, 104, 212-225. <http://dx.doi.org/10.1016/j.indmarman.2022.03.011>
- Mokady, R., Hertz, A., & Bermano, A. H. (2021). Clipcap: Clip prefix for image captioning. *arXiv preprint arXiv:2111.09734*.
- Nanne, A. J., Antheunis, M. L., Van Der Lee, C. G., Postma, E. O., Wubben, S., & Van Noort, G. (2020). The use of computer vision to analyze brand-related user generated image content. *Journal of Interactive Marketing*, 50(1), 156-167. <http://dx.doi.org/10.1016/j.intmar.2019.09.003>
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001), 2001.
- Philp, M., Jacobson, J., & Pancer, E. (2022). Predicting social media engagement with computer vision: An examination of food marketing on Instagram. *Journal of Business Research*, 149, 736-747. <http://dx.doi.org/10.1016/j.jbusres.2022.05.078>
- Syrdal, H. A., Myers, S., Sen, S., Woodroof, P. J., & McDowell, W. C. (2023). Influencer marketing and the growth of affiliates: The effects of language features on engagement behavior. *Journal of Business Research*, 163, 113875.
- Vedantam, R., Lawrence Zitnick, C., & Parikh, D. (2015). Cider: Consensus-based image description evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4566-4575). <http://dx.doi.org/10.1109/CVPR.2015.7299087>
- Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3156-3164). <http://dx.doi.org/10.1109/CVPR.2015.7298935>