Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: www.tandfonline.com/journals/hihc20

BeautyClicker: Modeling Viewers' Click Intent of the Cover Image to Support Drafts of Beauty Product Promotion Posts

Xueer Lin, Xia Chen & Zhenhui Peng

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2025.2531291

BeautyClicker: Modeling Viewers' Click Intent of the Cover Image to Support Drafts of Beauty Product Promotion Posts

Xueer Lin, Xia Chen and Zhenhui Peng

School of Artificial Intelligence, Sun Yat-sen University, Zhuhai, Guangdong, China

ABSTRACT
Promoting beauty products via posts on social media like RED has become an occupation for many people, and an attractive cover image can increase the click-through rate of these posts. However, selecting a cover image and writing the text for product promotion could be challenging, especially for novices in this occupation. In this paper, we propose a data-driven approach, with the best model's accuracy of 0.775, to assess whether viewers will click on the beauty product post in RED based on the content, color, and texture features of its cover image. With this model, we develop *BeautyClicker* to support drafts of beauty product promotion posts. A within-subjects study with novices (N = 32) and interviews with four beauty influencers demonstrate that *BeautyClicker* helps select cover images and improves the quality of posts. Lastly, we discuss insights into modeling viewers' click intent and supporting the creation of social media posts.

KEYWORDS
Beauty product promotion posts; click intent; computational assessment; AI-assisted content creation

CCS CONCEPTS
Human-centered computing; Human computer interaction (HCI); Interactive systems and tools

1. Introduction

Social media platforms, such as RED (Xiaohongshu in Chinese), TikTok (Douyin in Chinese), Weibo, Instagram, and X (Twitter), provide convenient ways for individuals to share their lives and discover intriguing and novel content (Barbosa dos Santos, 2022; Park et al., 2014). Within these platforms, content creators can document their lives through short videos, images, and text. They can also gain likes and bookmarks by sharing the experience of product utilization or recommending useful products for commercial purposes. In this paper, we focus on the influencers who actively share the experience of product utilization or recommending useful products for commercial purposes in these platforms (Alves De Castro & Carthy, 2021; Harrigan et al., 2021; Singh Tanwar et al., 2022). One such representative is beauty influencers, who actively share makeup techniques and promote beauty products (Abidin & Ots, 2016; Qin et al., 2024b; Zia et al., 2021). The recommendations made by beauty influencers have been shown to positively correlate with viewers' intention to purchase beauty products (Balqis et al., 2024; Chen & Dermawan, 2020; Ramadhan et al., 2025). Consequently, an increasing number of branded merchants are choosing to collaborate with beauty influencers to promote their products (Bakri, 2023), turning influencers into a primary or secondary occupation for many individuals.

A successful product promotion post needs to attract the viewers' attention via its cover listed along with the covers of other posts in the platform (Figure 1). A post's cover usually contains an image and a title, which dictate the viewers' first impression and influence their decisions to delve further into the post's contents, such as additional images, textual descriptions, or videos (Jalali & Papatla, 2016; Zhao et al., 2022). A high click-through rate for product promotion posts can boost the post's likes and bookmarks, leading to greater product visibility. This, in turn, helps the influencer gain more followers and attracts collaboration from advertisers (Bahtar & Muda, 2016; Chandrasekaran et al., 2019; Cheong & Morrison, 2008; Malik & Srinivasam, 2020). Traditionally, influencers can receive general guidance

CONTACT Zhenhui Peng pengzhh29@mail.sysu.edu.cn School of Artificial Intelligence, Sun Yat-sen University, Zhuhai, Guangdong, China

Supplemental data for this article can be accessed online at https://doi.org/10.1080/10447318.2025.2531291.

(a) The covers of beauty product promotion posts.

(b) The elements of the beauty product promotion post.

Figure 1. Example covers and content of beauty product promotion posts in RED.

from platforms about preparing the cover for their post. Take the RED as an example, the platform's "Creator Academy" prompts users that the cover images should pay attention to color matching, be high-quality, and contain keywords of the posts (Creator Academy, 2021). Besides, previous works have indicated that certain image elements, such as the facial attractiveness (Peng et al., 2020a; Xie et al., 2023) and clothing (Jin & Ryu, 2020), could affect the performance of the influencers' posts. However, it could still be difficult for beauty influencers, especially the novice ones who lack experience in creating the promotion posts, to implement such empirical suggestions. On the one hand, some elements (e.g., the products) in the cover images can not be changed by the influencers. On the other hand, the effect of one image on viewers' perceived impressions is usually the interplay of various visual features (Joo et al., 2014; Wei & Stillwell, 2017). Another way to facilitate the selection of a post's cover is enabling the creators to upload multiple covers, distributing them to a group of target audience for a short time (e.g., 12 hr), and giving feedback on the click-through rate of each cover. The RED platform internally experimented such A/B testing feature with a few influencers and did not luanch it now (Creator Academy, 2022). Possible reasons are that this A/B testing gives delayed feedback, and it involves a portion of the target viewers being exposed to images that may give wrong impressions. In brief, it would be beneficial for the beauty influencers to understand what contents could be shown in the covers of their product promotion posts and how the combination of these contents might affect viewers' click intentions on the posts before the publications, but few practical guidelines and real-time assessment on their specific covers are readily available.

The success of computational image assessment in other scenarios concerning user perception, such as brand personality (Wu et al., 2019) and first impressions of with images in online medical crowdfunding campaigns (Guo et al., 2022a), motivates us to explore the possibility of deriving a computational model of viewers' click intention on a beauty product promotion post based on its cover's features. Compared with other scenarios mentioned above, the beauty product promotion posts could emphasize more on the aesthetic preferences (BrydgesSjöholm, 2019) and integration of textual and visual features in the cover images (Creator Academy, 2021). To this end, we first collect 920 posts that promote female beauty products in RED,[1] a popular Chinese social media platform. We select posts using five beauty-related keywords to ensure a diverse range of content. These posts include both image-text and video formats. We avoid filtering posts by popularity to ensure a more comprehensive understanding of engagement levels, capturing a wide range of interaction data. According to data from April 2024, Xiaohongshu has 300 million active users and and over 80 million community contributors, with a male-to-female ratio of 3:7.[2] We then recruit 28 female RED users (age: Mean = 22.64, SD = 1.91) who are interested in beauty products to label whether each post would or would not be clicked based on its cover. With this annotated dataset of beauty product promotion posts' covers, we seek answers to the following research questions: RQ1) What are the relations between features of a cover and the viewers' click intention on the beauty product promotion posts? RQ2) Following RQ1, how could beauty influencers leverage the computational assessment of the post's cover to prepare their beauty product promotion posts? To address RQ1, we compute the 63 features as input in the binary click intent classification task, including the title on the cover as well as the color, texture, and content of the cover image. We train multiple models and examine combinations of different input features. The best one, Random Forest using ten features (e.g., the percentage of face area in the image and the percentage of orange pixels) of the cover image as input, achieves an accuracy of 0.7554. We invite six female RED users (age: Mean = 22, SD = 0.63) to make sense of how the features of the best model affect their intent to click on the cover of a post (Kwak et al., 2022). Our findings suggest that the percentage of face area and the color orange in the cover images could postively predict users' intention to click on beauty product promotion posts in RED. With this model for click intent prediction and interview findings with four female beauty influencers (age: Mean = 24.25, SD = 2.50), we develop ClickMe as a web app to explore RQ2. In ClickMe, users can upload a candidate cover image, get a predicted score on viewers' click intent, and get an assessment and LLM-generated suggestions on its visual features. We also customize the prompts that include detected text in the cover image to the LLM for generating a title and content of the beauty product promotion posts. Our within-subjects study with 32 female novice influencers (age: Mean = 21.32, SD = 0.95) indicates that compared to the baseline condition without our click intent prediction model, in the ClickMe condition, more participants were able to select the cover images of the experienced beauty influencers' choice. Besides, the outcome posts with ClickMe are perceived significantly more persuasive, trustworthy, and relevant to the product. Furthermore, in the expert interviews, four experienced beauty influencers (age: Mean = 24.25, SD = 2.50) highlight the values our model for helping them select a cover image. Lastly, we reflect on the insights of our computational model and user studies into modeling viewers' click intent on beauty product recommendation posts.

In summary, our contributions are three-folds:

- We contributes an interpretable computational model to predict viewers' intentions to click on a beauty product promotion post based on its cover image, highlighting the importance of features such as face area proportion and the area of the color orange in influencing viewers click intent.
- We propose a structured workflow that integrates our trained model to predict click intentions while leveraging a large language model to assist beauty influencers in crafting more effective beauty product promotion posts.
- We provide empirical understandings of how click intent prediction can help users draft social media posts via a user study of our proposed *BeautyClicker*.

2. Related work

2.1. Influencers on social media platforms

With the development of the Internet, an increasing number of people are sharing their lives and recommending products on social media. These individuals, who engage in the creation of content such as images and videos and disseminate it on social media platforms, are known as influencers and often receive likes and saves from their audience. Previous research has been conducted on influencers on social media platforms (Alves De Castro & Carthy, 2021; Harrigan et al., 2021; Singh Tanwar et al., 2022). For example, Harrigan et al. (2021) find that social media influencers can help spread information about product and service innovations, product launches, and marketing campaigns. The Human–Computer Interaction (HCI) community has conducted extensive research targeting the cohort of influencers

(Bakshy et al., 2011; Chou et al., 2023; Ma et al., 2023). For example, Ma et al. (2023) elucidate on how content creators interact with multiple platforms. This offers insights into designing enhanced creative support tools for cross-platform creators. Chou et al. (2023) delve into the reasons behind the appeal of social media influencers to users, revealing that an influencer's charm can be delineated into three categories: content, presentation, and closeness. Apart from the aforementioned investigation, HCI community researchers have also studied influencers in specific domains, exploring their work processes, and challenges, and providing design references for future work. For instance, Weber et al. (2021) conduct research on food influencers, highlighting challenges such as monetizing posts and using existing tools. It also proposes design considerations for relevant platforms and tools. Similarly, Zhou et al. (2023) conduct semi-structured interviews with mom vloggers. The research reveals that mom vloggers face challenges in balancing video production with parenting, thereby emphasizing the need for more convenient video production tools. Although previous studies have revealed numerous challenges faced by influencers, supportive work targeting this group remains relatively limited.

Our research focuses on beauty influencers, a substantial cohort present in the realm of social media. HCI scholars have engaged in the investigation of beauty content on social media platforms. For instance, Truong et al. (2021) autonomously generate hierarchical tutorials from makeup instructional videos, facilitating viewers in effectively navigating specific makeup procedures. Apart from cosmetic tutorials, beauty influencers on social media disseminate posts recommending beauty products. Our research predominantly centers on product promotion posts disseminated by beauty influencers. Current research within the HCI community predominantly consists of qualitative or quantitative studies focused on influencers, with a lack of models or tools specifically developed to address the challenges faced by bloggers and support this group. Our research endeavors to address this gap to a certain extent.

2.2. Product promotion on social media platforms

Promoting products on social media has become a prevalent strategy for brands to engage consumers and drive sales (Lindsey-Mullikin & Borin, 2017; Mangold & Faulds, 2009; Singh, 2021; Tafesse & Wood, 2021). Research in Human–Computer Interaction (HCI) illustrates that content, presentation, and closeness are why people favor social media over traditional media (Chou et al., 2023). Prior research in the marketing area has predominantly focused on how influencers leverage their social media platforms to establish trust with their followers (Kim & Kim, 2021; Lou & Yuan, 2019) and implement effective content strategies for product promotion (Haenlein et al., 2020). HCI researchers summarize eight subtasks of the content creator's process (Weber et al., 2021), which are monetization, inspiration, meal preparation, photography set construction, taking pictures, editing, posting, and monitoring. Within this paper, we focus on the "Editing stage." In this stage, the selection of images is equally critical to crafting the text of a post, as they directly influence the effectiveness of the promotional message (Olivier et al., 1999). However, selecting appropriate images can be challenging, as they must not only be visually appealing (Azimi et al., 2012; Bazi et al., 2023). but also align with the brand's identity and the product's attributes (Dennhardt, 2014; Janssen et al., 2022). Azimi et al. (2012) describe the impact of the visual appearance of advertising materials on users' propensity to respond and point out that the visual features in the materials could predict click-through rates. Moreover, drafting the text in the post also requires careful consideration of language and tone (Hübner Barcelos et al., 2018; Lee & Theokary, 2021) and its relevance with the image, as Olivier et al. (1999) highlights that participants' perceived advertising effectiveness of tourism photographs is fundamentally affected by the text language. Our work is motivated by the unique appeal of social media influencers for product promotion. We focus on modeling viewers' click intent on the product promotion posts and studying how such models could facilitate influencers in the post-editing stage.

2.3. Computational models for understanding user perceptions of posts and images

Existing HCI work has explored a variety of computational models to understand user perception of social media posts, especially the textual posts, *e.g.*, identify promises and supporting points (Liu et al., 2022), extract helpful critiques (Peng et al., 2024), and predict the sought social support in the posts

(Peng et al., 2021). For example, Peng et al. (2024) develop DesignQuizzer to assist beginners in learning visual design based on the design examples and critiques in online communities. They fine-tuned models like "T5-small" and "RoBerta-based" to extract meaningful feedback, classify sentences, and detect visual design keywords (*e.g.,* color, space) from the textual comments. Research on user perceptions of visual content is also growing (Shin et al., 2020; Zhang et al., 2021a), *e.g.,* predicting perceived brand personality of mobile app UIs (Wu et al., 2019) and first impressions of images in online medical crowdfunding campaigns (OMCCs) (Guo et al., 2022b). Specifically, Guo et al. (2022b) train machine learning models like random forests based on 900 labeled images to assess whether an OMCC image conveys appropriate first impressions (*i.e.,* empathy, credibility, justice, impact, and attractiveness), which are positively related to viewers' donation intention. Furthermore, recent research has begun to model user perceptions of text-image posts (Koh & Cui, 2022; Li et al., 2024; Oh et al., 2020). For instance, Li et al. (2024) explore text-image posts in grief support communities and use regression analyses to show that a post is also likely to get more social support if its text describes the visible content or tells a story depicted in the image.

This line of previous work has paved the way to modeling viewers' click intent on beauty product promotion posts based on their cover image and title. In addition, in the field of predicting user click intent, Cui et al. (2024) analyzed emotions in YouTube video titles and cover image to predict view counts. Their research found that strong emotions in cover image lead to more views, and titles with positive emotions generate more views than those with negative emotions. The study by Lee et al. (2023) explored the relationship between visual attributes (*e.g.,* celebrity endorsement, colorfulness) of YouTube video thumbnails and view counts, proposing a novel approach based on visual cue classification using the elaboration likelihood model. In contrast, our study focuses on the unique challenges of beauty product promotion posts, which require understanding the interplay of aesthetic appeal and image content, making the task more complex and nuanced. Furthermore, by building computational models, we can provide interpretability for the factors that influence user click intent. This insight helps influencers identify which features, such as visual elements and image content characteristics, are most effective in capturing their audience's attention. As a result, it can reduce hesitation among novice influencers and assist them in making more informed decisions when creating posts.

2.4. Intelligent assistants for content co-creation in social media

In the field of content creation, various studies and tools have emerged to assist users in drafting high-quality posts or promoting a sense of participation. For instance, Peng et al. (2020b) introduce a technological prototype called MepsBot, which provides users with real-time writing assistance by evaluating their text and recommending examples to enhance the quality of their comments in online mental health communities. Besides, Liu et al. (2023) propose a tool with comment navigation and a chatbot to encourage lurkers to contribute answers in question-answering communities. Recent advancements in the application of large language models (LLMs) have significantly impacted writing assistants across various domains, such as argumentative writing (Zhang et al., 2023) and tweetorial writing for scientific communication (Gero et al., 2022). For example, Zhang et al. (2023) introduce visual programming elements in an LLM-powered tool to assist users in argumentative writing. Ding et al. (2023) explore how to maximize the use of LLMs for writing news headlines and analyzed the impact of interacting with these models on people's trust during the writing process. Additionally, Existing research has demonstrated that LLM-based chatbots can effectively assist users in content creation. For instance, the CharacterMeet tool proposed by Qin et al. (2024a) supports writers in iteratively developing character profiles through conversations with the chatbot. However, there is a lack of understanding of how beauty influencers view the writing support tools for drafting product promotion posts. Moreover, few works have explored supporting the editing of posts with respect to image selection, which is important for a product promotion post. Beauty product promotion posts present unique challenges, requiring a balance between aesthetic appeal and functional attractiveness to capture audience interest. Moreover, our tool aims to integrate multi-modal image features, which consider both visual aesthetics and image content features. It focuses on supporting novice creators in crafting compelling content. In this work, we develop and evaluate *BeautyClicker* that leverages our click intent prediction model to support the selection of the cover image for beauty product promotion posts. In addition, the tool integrates an LLM-based creative assistant to support influencers in generating post drafts and introduces an LLM-based chatbot to help optimize post content.

3. Modeling viewers' click intention on beauty product promotion posts

Figure 2 depicts our research flow. In this section, we describe how we model viewers' click intention on beauty product promotion posts.

3.1. Research site and dataset

We use data from the RED platform, one of the largest social media platforms in China. To identify beauty product promotion posts, we first choose five beauty-related Chinese keywords in RED, all of which have garnered viewers exceeding ten million. These keywords, after being translated into English, are "Beauty Product Promotion," "My Beauty Favorites," "Affordable Cosmetics," "Annual Makeup Favorites," and "Noteworthy Makeup Products." For each keyword, we retrieve 226 posts using a web crawler,[3] resulting in a total of 1130 image-text or video-text posts. Each post includes an image, a title, and a creation timestamp. Furthermore, we manually filter out posts unrelated to female-oriented beauty product promotion or were repeated. After these steps, we obtain a dataset of 920 posts. This number of collected posts could be sufficient for building a classification model, as does in previous HCI work (Guo et al., 2022b; Peng et al., 2020b). The timestamps of these posts span from March 2022 to February 2024. We avoid filtering posts by popularity, intending to obtain posts with various levels of viewers' engagement. Overall, the numbers of likes under the collected posts range from 0 to 529903 ($Mean = 12945.23$), and the numbers of comments range from 0 and 66790 ($Mean = 339.63$).

We carry out an annotation task to label each post as either "will be clicked by viewers" or "will not be clicked by viewers" based on its cover. Due to privacy concerns, we are unable to access the views metric provided by RED, as it is only visible to the post owner. As a result, we rely on this annotation task instead of click data to make the determination. We recruit 28 female RED users who are interested in beauty and makeup as annotators via word-of-mouth in the local university. Their ages range from twenty-one to twenty-seven years (Mean = 22.64, SD = 1.91). Among them, twenty are university students majoring in fields such as Artificial Intelligence, Computer Science and Technology, and Materials Chemistry, while eight are full-time employees. To understand their familiarity with RED beauty content, we also collect information about their browsing frequency. Eighteen participants report that they browse beauty-related posts on RED on a daily basis, eight participants browse such content two to six times per week, and two participants report browsing once per week. In particular, we

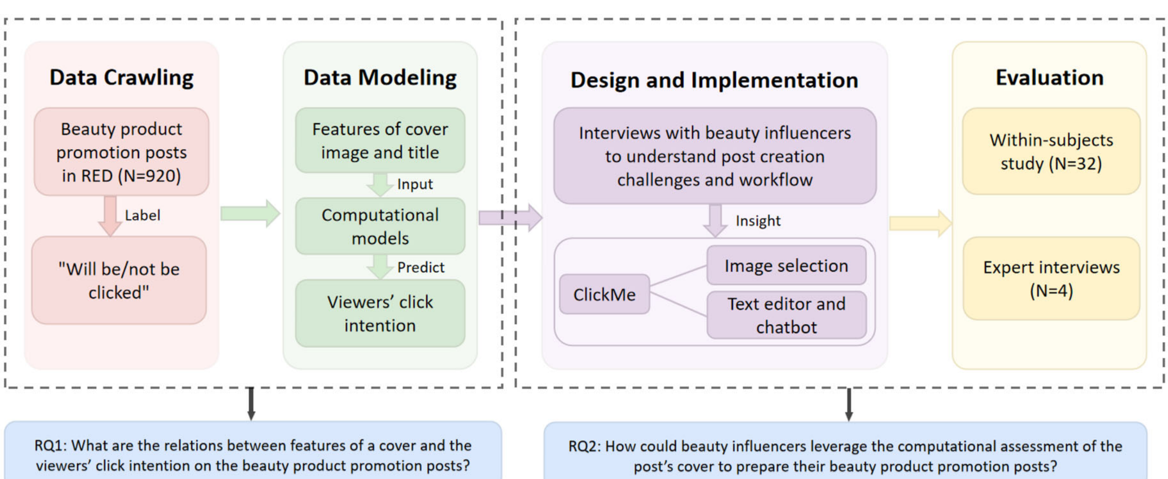

Figure 2. Research flow of modeling viewers' click intent of the cover image to support drafts of beauty product promotion posts.

recruit 28 annotators following (Guo et al., 2022a) to strike a balance between annotation quality and diversity of opinions. This number allowed us to collect a broad range of subjective judgments while keeping the task manageable. Using multiple annotators also helps reduce the influence of individual preference on the final labels. We carry out the annotation task on the Wenjuanxing,[4] a popular survey platform in China. Following practices in (Guo et al., 2022b; Peng et al., 2020b), we assign each post to five annotators. Specifically, the 920 posts are ramdomly assigned into 28 questionnaires. Each questionnaire contains 32 or 33 posts. Each annotator is randomly assigned five different questionnaires. The design of the questionnaire simulates the interface of the RED homepage, which lists a set of posts with their covers one by one. We prompt the annotators to emulate the use of the RED interface and select the posts that they would click on at the first glance in their assigned questionnaire. On average, each annotator spends 126.2 s (SD = 34.90 s) in our task and receives 5 RMB as compensation. We determine the label for each post based on majority vote, *i.e.,* if a post is selected by three or more than three annotators, its label is "will be clicked by viewers." The annotation results of these posts, specifically the counts of how each post is labeled as "will be clicked by viewers" or "will not be clicked by viewers," are shown in the Table 1. In the end, 204 posts are categorized as "will be clicked by viewers," and the rest 716 posts are categorized "will not be clicked by viewers."

3.2. Predictive features

The goal of our intended model is to predict viewers' click intent based on the features of its cover, including a cover image and a title. Inspired by related work (*e.g.* Guo et al., 2022b), we select a set of possible predictive features and categorize them into four types: title-based, color-based, texture-based, and content-based. The first type of features pertains to the title, while the subsequent three types are concerned with the cover image. All these features are summarized in Table 2 and described below.

3.2.1. Title-based feature

Kolmogorova et al. (2021) suggest that the emotional orientation conveyed by titles, encompassing positive, negative, or neutral tones, affects people's memory retention of the text. Besides, Nuno Emanuel Branquinho Moutinho Marques de Paiva (2018) indicate that on e-commerce platforms, the use of emoji affects people's informal perception of language, adding an element of fun to the text and thereby influencing the recommendation of services. Existing work also indicates that the inclusion of precise numbers within text can impact user behavior. For instance, Wadhwa and Zhang (2015) suggest that rounded numbers (*i.e.,* numbers ending in a whole number) can affect consumers' purchase intentions, particularly when product information evokes positive emotions. Therefore, we include the feature about whether titles contain numbers, considering both Chinese numerals and Arabic figures.

3.2.2. Color-based feature

Previous studies indicate that the color of an image can affect individual perception of it. For example, Yu et al. (2020) discover that the hue and brightness of an image can influence the popularity of posts on Instagram, while Osgood (1957) suggest that the saturation of an image can impact the emotions of the viewers. We also integrate semantic color area distribution, determining the proportion of pixels for eleven primary colors based on their corresponding HSV channel range values (Van De Weijer et al., 2007). Equivalently, the HSV area distribution measures the spatial coverage of pixels varying in brightness, saturation, and hue (Van De Weijer et al., 2007). These two color distribution features offer distinct perspectives for an image's color composition, and they are also utilized in the task of emotion classification (Machajdik & Hanbury, 2010). Furthermore, vibrant and saturated colors are preferred in many cultures because they are often associated with positive qualities such as beauty, youth, and

Table 1. The ratio of "will be clicked by viewers" to "will not be clicked by viewers" and the corresponding post counts.

"will be clicked": "will not be clicked"	0:5	1:4	2:3	3:2	4:1	5:0
Number of Posts	177	300	239	144	55	5
Total Post Count		716			204	

Table 2. The feature set of the image and title of a post cover for predicting whether viewers intend to click on it.

Feature set	Feature name	Number	Description	Source
Title-based	Text sentiment	1	Emotional orientation (positive, negative, or neutral)	Kolmogorova et al. (2021)
	Emoji number	1	The number of emoji in the title	Nuno Emanuel Branquinho Moutinho Marques de Paiva (2018)
	Number	1	Whether there are any numerals in the title	Wadhwa and Zhang (2015)
Color-based	**HSV statistics**	6	**Mean of brightness** and saturation, hue distribution in 4 dimensions	Osgood (1957) and Yu et al. (2020)
	Semantic color area distribution	11	Percentage of black, blue, brown, green, gray, **orange**, pink, purple, red, white, **yellow** pixels	Van De Weijer et al. (2007)
	HSV area distribution	10	Based on Wang features, area of brightness (very low, low, middle, high, very high), area of saturation (high, middle, low), area of hue (warm, cool)	Van De Weijer et al. (2007)
Texture-based	Tamura	3	Three tamura features (coarseness, contrast, directionality)	Liu et al. (2015)
	Wavelet	12	**Wavelet textures(spatial smoothness/ graininess) in three levels on each HSV channel**, sum of all levels in each channel	Tan et al. (2018)
	GLCM	12	**Contrast, correlation, energy, homogeneity for three HSV channels**	Haralock and Shapiro (1991)
Content-based	Face	1	Whether there are faces in the image	Kim et al. (2016) and Steele et al. (2009)
	Face area percentage	1	**The percentage of face area in the image**	Bakhshi et al. (2014), Olivier et al. (1999) and Zhang et al. (2021b)
	Face number	1	The number of face in the image	Ren et al. (2020)
	Text in image	1	Whether there are texts in the image	Li et al. (2016)
	Text percentage	1	The percentage of text area in the image	Pieters and Wedel (2004)

These features are used to train different click intent classifiers. The features in bold are the input to the classifier with the best performance.

vitality. This is especially true in beauty-related contexts, where these colors are used to convey energy, health, and appeal (Elliot & Maier, 2015). For example, bright and rich hues are commonly used in beauty product packaging and advertisements to convey energy and appeal, aligning with societal ideals of attractiveness and health (Labrecque & Milne, 2012). In our study, the color-related features include HSV statistics, semantic color area distribution, and HSV area distribution. They provide a detailed analysis of an image's color composition, focusing on brightness, saturation, and hue. These features help in understanding how color influences viewer perception.

3.2.3. Texture-based feature

Previous research reveals that the textural features of images play a role in shaping the impressions of human faces and contribute to the assessment of image appeal (Russell et al., 2006). In the case of beauty-related content, texture can influence perceptions of the skin quality, the smoothness of product application, and the overall aesthetic appeal of beauty products (Humphrey et al., 2021). The tamura texture of images is utilized to capture the emotional perception of visual textures (Liu et al., 2015). The tamura features are thus essential for understanding how viewers perceive product aesthetics and quality (Porcheron et al., 2014), particularly in contexts such as makeup tutorials or skincare products. We use features such as coarseness, contrast, and directionality, which can grasp advanced perceptual characteristics of texture and are widely applied in the realm of visual art appreciation (Liu et al., 2015). Wavelet texture features have been used in previous studies to investigate facial attractiveness (Tan et al., 2018). Furthermore, wavelet texture features in the HSV color space are represented by three levels of wavelet textures (spatial smoothness/granularity) in each of the three channels, along with the sum of wavelet textures across all levels in each channel. Additionally, GLCM features are commonly employed in texture analysis (Haralock & Shapiro, 1991). For each image, we compute the contrast, correlation, energy, and homogeneity corresponding to the three channels of HSV by the GLCM function.

3.2.4. Content-based feature

Existing works indicate that the content of an image influences the initial impression of the viewers (Guo et al., 2022b; Kim et al., 2016; Steele et al., 2009). This study extracts two types of image content features: facial features in the image, and textual characteristics of the image. We contemplate the presence of faces in an image, the number of faces (Ren et al., 2020), and the proportion of the image area occupied by the faces (Olivier et al., 1999; Zhang et al., 2021b). Prior studies demonstrate that text on photographs is capable of capturing the visual attention of viewers (Li et al., 2016; Pieters & Wedel, 2004). Regarding textual characteristics, our investigation focuses on the existence of textual elements and the proportion of the image's area they occupy. We compute content-based features with the Baidu AI Cloud API.[5] The API can analyze face and text attributes for each image with reliable accuracy. After invoking the API to identify faces and text in images, we also conduct a sampling inspection of the identification results to mitigate instances of excessive recognition and misidentification.

3.3. Computational models

We follow the feature-selection strategies suggested by previous work (Guyon & Elisseeff, 2003; Louppe et al., 2013) to train and compare multiple computational models. Specifically, we first train a set of common models using all features extracted above as input. Then, for the best-perform model (random forest in our case), we examine its performance using different feature sets as input, including the title-based, color-based, texture-based, content-based, and top-k important features.

3.3.1. Common models using all features as input

Based on the extracted predictive features of a post's cover image and title, we train a series of computational models to predict viewers' click intent on this post. We formulate the click intent prediction task as a binary classification task, following (Guo et al., 2022b). Each cover has a label, which is "will be clicked by viewers" or "will not be clicked by viewers."

We adopt five commonly used machine learning algorithms, including K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Multi-Layer Perceptron classifier (MLP), Decision Tree, and Random Forest for classification. Previous research has utilized these algorithms to model intent or classify community data in various contexts (Guo et al., 2022b; Joo et al., 2014; Wu et al., 2019). These machine learning models take the extracted features in Table 2 as input, and are implemented with scikit-learn packages (Pedregosa et al., 2011). We have also constructed a multimodal deep learning neural network for training, which utilizes the pre-trained BERT model and ResNet-34 with Pytorch (Paszke et al., 2019) to encode text and images respectively. Specially, the text comprises both the title and the text in the image. The encoded vectors are concatenated as input to a fully connected layer with the dimension of 1664×2 for predicting viewers' click intent on the post.

To evaluate the performance of the trained classifiers, we employ precision, recall, and the F1 score (Yacouby & Axman, 2020) as standard metrics. We randomly divide the data into a training set (80%) and a testing set (20%). Initially, we use ten-fold cross-validation on the training set to ascertain the model's hyper-parameters. For the KNN model, the number of nearest neighbors is set to 10. The SVC uses a Sigmoid kernel, with the regularization parameter set to 0.1. The MLP adopts a single hidden layer powered by the RELU activation function, with the number of neurons in this hidden layer being 100. Regarding the decision tree model, the maximum depth of the tree is set to 5. For the random forest model, the number of trees is fixed at 80, with a maximum depth of 8 for each tree. For deep learning neural networks, we initially partition 25% of the training data to serve as the validation set and employ early stopping on the validation set to select the best hyper-parameters (Caruana et al., 2000).

Table 4 summarizes the precision, recall, and F1-score achieved by the six models mentioned above. When compared to baseline models, such as KNN (precision $= 0.7283$) and SVC (precision $= 0.7174$), the Multimodal Deep Learning Neural Network (precision $= 0.7663$) demonstrates a significant improvement in precision. However, it performs poorly in recall (0.4038), suggesting its challenge in identifying positive instances. On the other hand, the Random Forest (RF) model achieves the highest recall (0.5501) and F1-score (0.5221) in this task, outperforming the baseline models. This also indicates the efficacy of the features we have extracted for training a click intent classifier given around 1000 label data points. Taking all features into account, the RF model emerges as the best performance with the highest recall (0.5501) and F1 (0.5221) in this classification task, despite its slightly lower accuracy compared to Multimodal Deep Learning Neural Network. One plausible reason for the effectiveness of the RF model is its ability to ensemble several decision trees, making it well-suited for tasks involving complex feature interactions and reducing sensitivity to hyperparameters in classification tasks with multiple features (Ali et al., 2012). Another reason is that Random Forest (RF) models are often more efficient with smaller datasets, whereas deep learning models typically require larger datasets to perform effectively. This finding aligns with previous studies, such as Guo et al. (2022b) showing that the RF performs the best (accuracy = 0.733) in predicting the first impressions of images in online medical crowdfunding campaigns and (Zhu, 2020) showing that the RF performs the best (0.78) in assessing the percerived brand personality of mobile app UIs.

3.3.2. Random forest using different feature sets as input

We also train RF models, the best-performing model when using all features as input, with different feature sets (Table 2) to understand their contribution to predicting viewers' click intent. The performance of these RF models is shown in Table 4. We can find that the RF model trained by texture-based features achieves the highest performance of precision (0.7337), followed by those employing color-based features (0.7174) and content-based features (0.6739). The one applying title-based features (0.6196) comes last. This outcome emphasizes the importance of texture-based features in predicting viewers' click intent of the cover image (Liu et al., 2015). However, the best performance is achieved by combining these different types of features into the model. This demonstrates the importance of taking into account features of a post's cover from multiple perspectives to predict viewers' click intent.

Concurrently, for the best-performing RF model using all features as input, we conduct Gini importance analysis (Breiman et al., 1998) to calculate the relative importance of each feature. Table 3 displays the ten most significant features within the RF model along with their respective feature importance values. The feature of face area percentage has the highest overall importance score in predicting viewers' click intent of RED covers. This also suggests that it is crucial to pay attention to the proportion of the face area in images when creating or selecting cover images for beauty product promotion posts. In addition, certain color-based features of images affect the performance of the classifiers, such as the area of color orange and the average of brightness. According to preceding study, the appeal of images in travel-related posts on Instagram is associated with the image's brightness and saturation (Yu et al., 2020). Especially, the color orange particularly engages viewers' interest in the content of the image, while audiences also show a preference for images that are both brighter and more saturated (Yu et al., 2020).

To further advance the performance of RF model, we reference strategies employed by predecessors (Guyon & Elisseeff, 2003; Louppe et al., 2013) and select the top ten, fifteen, and twenty features based on their importance scores. As shown in Table 4, the RF model using the top ten features has the best performance. Compared to the RF model with all features, the RF model with top ten important

Table 3. The top-10 important features for predicting viewers' click intent in the Random Forest that uses all features (Table 2) of the cover image and title as input.

Feature	Importance value	The name in *BeautyClicker*
Area of face	0.0440	Face Area Percentage
Area of color orange	0.0277	Orange Element Proportion
GlcmS.energy	0.0240	Color Vibrancy
GlcmV.contrast	0.0234	Contrast of light and shadow
The third level of wavelet - texture value of V channel	0.0233	Intensity of Macroscopic Texture
Brightness average	0.0223	Brightness Average
The second level of wavelet - texture value of V channel	0.0220	Comparison of Intricate Textural Details
GlcmS.correlation	0.0219	Color Harmony
GlcmH.correlation	0.0216	Tone Continuity
Area of color yellow	0.0212	Yellow Element Proportion

The name of each feature in *BeautyClicker* is also listed in this table, with references in Table 2.

Table 4. The overall performance of classifiers.

Input features	Model	Precision	Recall	F1
All features	KNN	0.7283	0.5251	0.4734
	SVC	0.7174	0.5000	0.4177
	MLP	0.6630	0.4854	0.4531
	Decision Tree	0.6957	0.5306	0.4881
	Random Forest	0.7391	**0.5501**	**0.5221**
	Multimodal Deep Learning Neural Network	**0.7663**	0.4038	0.4941
Different types of features	Random Forest (title-based features)	0.6196	0.4784	0.4689
	Random Forest (color-based features)	0.7174	0.5117	0.4524
	Random Forest (texture-based features)	**0.7337**	**0.5580**	**0.5422**
	Random Forest (content-based features)	0.6739	0.4988	0.4708
Important features	Random Forest (top ten important features)	**0.7554**	**0.5906**	**0.5894**
	Random Forest (top fifteen important features)	0.7337	0.5463	0.5186
	Random Forest (top twenty important features)	0.7283	0.5367	0.5022

features has superior precision (0.7554), recall (0.5906), and F1 (0.5894). The enhancement in the F1 score particularly highlights the boosted capability of the advanced RF model in identifying positive cases. Besides, the RF model with top ten important features outperforms the RF model trained on separate feature sets, which suggests that predicting viewers' click intent on product promotion posts in RED cannot rely merely on singular categories of features but requires a comprehensive analysis combining distinct features across various categories. We notice that the top ten features used in the advanced RF model are all from the cover image of a post. Therefore, in the later design and development of *BeautyClicker*, we consider predicting viewers' click intent on a post only based on its cover image, though viewers can also see its title on the RED homepage.

3.3.3. Understanding the features used for predicting viewers' click intent with six RED users

While the feature analyses of the computational models above indicate the importance of the image's features on predicting viewers' click-intent of the post, we need to investigate if human viewers would also consider these features as important. For this purpose, we conduct interviews with six female RED users (U1-6, age: Mean = 22, SD = 0.63), who have experience browsing such posts. We seek to explore how these users perceive the features and whether they align with the features selected by our model. Among them, four participants browse beauty-related posts on RED daily, while two participants browse such posts 2-6 times per week. This step allows us to assess the relevance of these features in predicting click intent and helps to refine our approach in future work. We acknowledge that the small sample size (N = 6) limits the generalizability of the findings and that there may be a degree of confirmation bias, as participants were asked to reflect on features our model had already highlighted. These limitations will be addressed in future studies with larger and more diverse samples. First, we inquire about their perspectives on factors influencing viewers' click intent on beauty product promotion posts. They all regard the cover image as a critical element in affecting viewers' click intent, which reflects the validity of our categorization of features. Like U2 states, *"Vivid colors and bright imagery in cover images are elements that capture my attention and stimulate me to click on a post when browsing the homepage."* Four individuals (U2, U3, U4, U5) discuss how they perceive the title of a post. For instance, U2 and U4 mention that titles could resonate with their personal needs, and a word like "budget-friendly" would increase their intentions of clicking on the post. Moreover, U1, U2, and U3 highlight the influence of product branding on their intent to click the posts. For instance, U2 remark, *"If the promoted product is a brand with which he is familiar, he will click on it."* Additionally, both U1 and U5 mention that if the publisher is a blogger they favor, this too would enhance the likelihood of their click. In particular, some users also mention encountering unexpected perspectives. U3 mentions that she is sometimes attracted to cover images with elements of curiosity, prompting her to click and explore the content. For instance, intentionally created "unattractive" makeup-free photos can spark her interest. In addition, U5 also expresses that after viewing too many polished cover images, those rougher images created by ordinary people seem more real and resonate more with her. Our models could not take into the account of such user-related features due to the lack of access to a larger amount of RED data. Nevertheless, these results indicate a promising future direction to incorporate the user features for predicting viewers' click intent on a post.

Specifically, we query the six RED users about their views on the top ten features (Table 3), particularly the top two features, area of face and area of color orange, of a cover image used in the advanced RF model. We illustrate these features with both positive and negative instances of cover images from our database and invite the RED users to articulate their perceptions. They reach a consensus about the crucial influence of these two factors. U1 notes that having an appropriate facial area in the image to demonstrate the product's effect would make her more likely to click on it than just a plain product display. U2 states, *"To keep the image looking good, the size of the face in it should not be too big or too small."* Regarding the area of color orange, U1, U2, and U6 mention that vibrant colors like orange align with the aesthetic preferences of young women on cosmetic products, capturing their attention more effectively. U5 remarks, *"The cover image featuring the area of color orange excellently conveys a sense of softness and credibility."* The comprehension exhibited by these participants validates the efficacy of the chosen features within the model.

4. Design and implementation of *BeautyClicker*

With the predictor and understanding of viewers' click intention on the posts, we conduct semi-structured interviews with four beauty influencers to design *BeautyClicker* in order to assist them in editing product promotion posts.

4.1. Interviews with beauty influencers

4.1.1. Procedure

To address RQ2 about how could beauty influencers leverage the computational assessment of the post's cover to prepare their beauty product promotion posts, we develop and evaluate a technical prototype *BeautyClicker* that incorporates our our click intent prediction model. In this section, we first describe the interviews with beauty influencers to understand the practices and challenges of preparing a cover for its post, which inform the design of *BeautyClicker*. Then, we present the implementation of *BeautyClicker*.

We recruit four female beauty influencers (I1-4, age: Mean = 24.25, SD = 2.50) via word of mouth. I1, I3, and I4, who have more than 10,000 followers, have been beauty influencers and promoted products in RED for over two years. Specially, I1 has more than 80,000 followers. The rest participant, I2, is a novice with three months of experience as a beauty influencer and garners 949 followers. Among these influencers, two are full-time content creators, while the other two are university students. During the interview with each participant, we discuss the processes involved in editing product promotion posts (*e.g.,* "What is the process for editing posts when you want to recommend a beauty product?") and the challenges encountered during selecting the cover image (e.g., "Do you often struggle with choosing a cover image?," "Do you possess insightful reflections when selecting a cover image?"). Specifically, we present our click intent classification model of a post's cover image and seek participants' opinions on it. We also investigate their perspectives regarding LLMs such as ChatGPT in the process of editing titles and content. Each interview lasts around 40 min and each participant receives 100 RMB as compensation.

4.1.2. Findings

All four beauty influencers value the potential of our click intent prediction model for supporting the drafts of product promotion posts. They share their practices, challenges, and expected support when editing a beauty product promotion post, which is summarized below.

4.1.2.1. Select a cover image. All participants acknowledge that the choice of a cover image is crucial for achieving high click-through rates for their posts. They share their perspectives on these features of an image inputted into our model. For example, I1 notes that the inclusion of human faces can reduce the perceived marketing intention and make it easier for fans to remember the content in the post. I1 also emphasizes the necessity of using text in the cover image to encapsulate the key points of the post. I2 reveals that she imitates the cover designs of popular posts, including their composition, shooting angles, and color schemes. I3 describes that she prefers a cover image with a cohesive aesthetic, as it creates a visually pleasing first impression. In addition, both I1 and I3 mention that for influencers with many years of experience, their cover images tend to develop a specific style, which can make their work more recognizable. Besides, three influencers (I1, I2, I4) express their hesitation when selecting cover images. For instance, I1 remarks, *"It is quite challenging to create a cover that both conveys the core message of the post and retains user interest."* All beauty influencers are positive that our click intent prediction model would reduce their hesitation in selecting cover images. However, they indicate that the model's extracted features of an image should be more understandable. Apart from a predicted score, they expect guidance on how to improve their posts and would like to check good and bad examples of cover images for promoting beauty products.

4.1.2.2. Compose the title and content. All influencers further indicate that *"creating the title and content of a post that attract viewers' attention and keep them engaged is a time-consuming and energy-intensive process,"* as I2 states. Although there are many hot product promotion posts available online, influencers need to avoid copyright issues and instances of plagiarism. Three participants (I1, I2, I4) have experiences in using AI writing tools powered by LLMs to edit product promotion posts, but the experiences are not satisfactory. For instance, I1 recalls, *"the tools could not understand the specific steps, content, and professional terminology in beauty makeup. The output is stiff, is not detailed or professional enough, and requires a major revision."* Our beauty influencers expect an AI writing tool that has pre-set prompts and custom input fields specialized for generating beauty product promotion posts.

4.2. User scenario and interface of BeautyClicker

The findings from the formative study with four beauty influencers inform the design of *BeautyClicker*, an intelligent assistant powered by our click intent prediction model and LLM GPT-4 for supporting drafts of beauty recommendation posts. To walk through the user interface of *BeautyClicker* (Figure 3), we describe Mary, a female beauty influencer whose goal is to publish a post (also called note in RED) that promotes a liquid makeup product. She has prepared a set of candidate cover images and has formal descriptions about the product's features provided by the company of the product at hand. Now she uses *BeautyClicker* to draft the product recommendation post.

To begin with, Mary uploads candidate cover images in the top-left panel of *BeautyClicker*'s interface (Figure 3 A1) and clicks the "Upload for parsing" button, which will return the predicted score under each image indicating its possibility of being clicked by viewers. Mary selects the image with the highest probability, after which she can see the panel that analyzes this image (B1) and see how it looks in the published post (C1). In the image analysis panel (B1), Mary can view the key features of the selected image, the value of each feature, and whether the value is lower or higher than the average value of positive/negative cases in our collected dataset. A positive case is a cover image that has a label of being clicked in our dataset, while a negative case is a cover image that has not been clicked. With an interest in the "Face Area Percentage" feature, Mary clicks it and sees the explanation of it and suggestions for improvement. She agrees that "a suitable face area ratio helps improve the click-through rate." Mary is also interested in the "Orange Element Proportion" feature of the image. This time, she selects this feature in the drop-down menu and clicks the "Filter Cases by Features" button in the image example panel (A2). She views multiple representative positive and negative cases in our dataset, which further enhances her understanding of how this feature is represented in the cover images. After checking image analyses of multiple candidate images, Mary is confident in selecting the one with the highest score on viewers' click intent as the cover image.

With the selected image, Mary clicks the "Text Editor" (Figure 3 B2), which will replace the image example panel with a chatbot (B3) that helps the user draft the title and content of her post. Mary can type down the topic of the note, product features and selling points, and anticipated user's feelings of the product based on the given product description from its company and her experience with the product. After that, she clicks the "Generate Title" and "Generate Content" respectively to get the generated title and content of her post. Mary can also ask the chatbot (B3) anything at any time, *e.g.*, about an example description of a product's features and functions with emojis. Mary is quite satisfied with

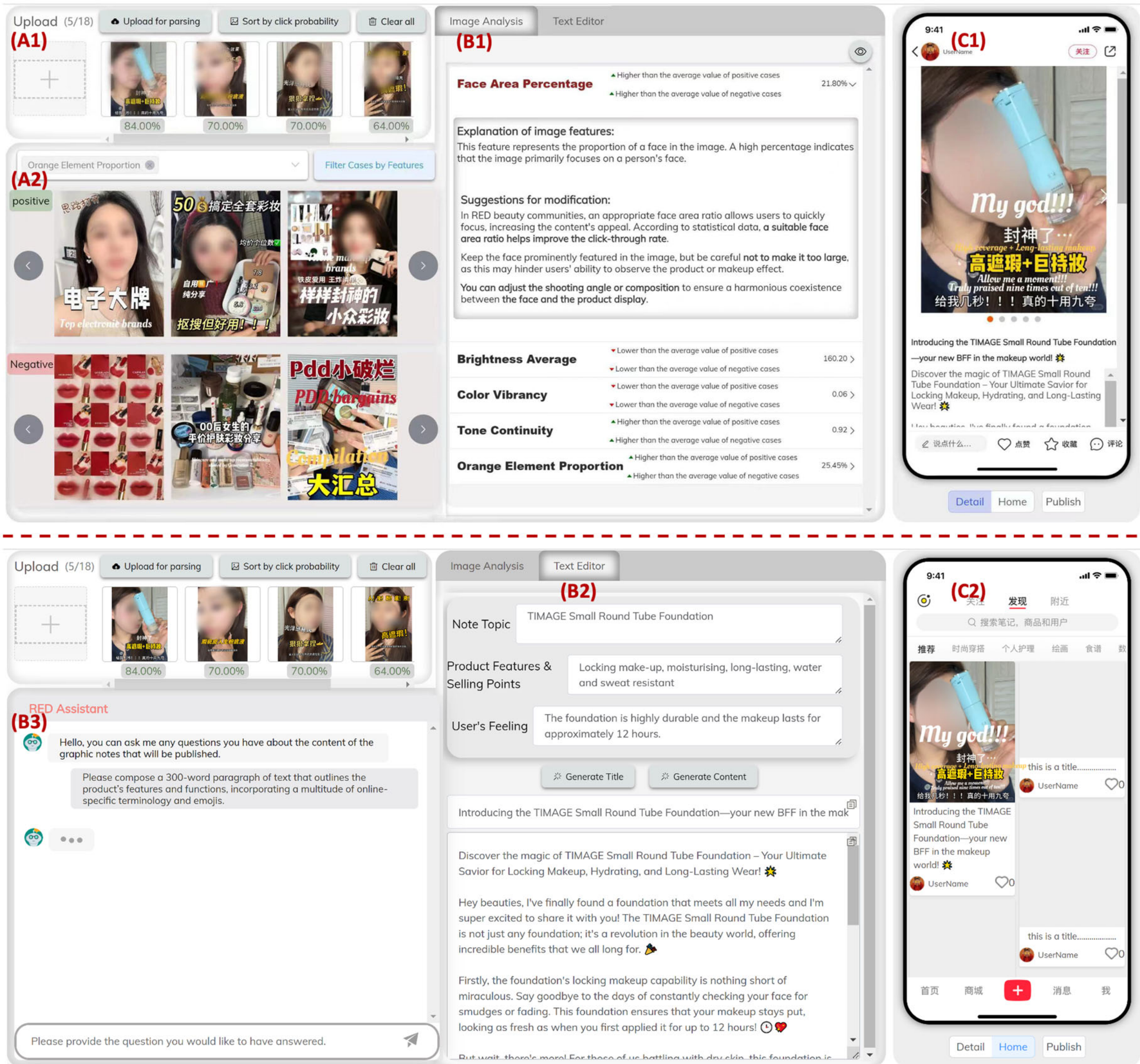

Figure 3. Interface of *BeautyClicker*. (A1) Upload candidate cover images panel. (A2) Image example panel. (B1) Image analysis panel (B2) text Editor. (B3) Chatbot powered by GPT-4. (C1) Preview panel of the detail page. (C2) Preview panel of the home page.

the generated title and previews it together with the cover image on the home page of RED (C2). However, she has her own thoughts on the content and adds it to the text editor. Finally, she is satisfied and clicks "Publish" in the bottom-right part of *BeautyClicker*'s interface to finish her post promoting a liquid makeup product.

4.3. Implementation

BeautyClicker is designed as a web application, with the frontend implemented using the Vue 2 framework and the backend developed with Django. We have chosen to utilize the OpenAI GPT-4-0125-preview as the large language model (LLM) to generate content within *BeautyClicker*. This decision balanced the trade-offs of cost and performance. Given the same prompt template and input, our research team empirically compared the outcome of GPT-4-0125-preview and that of GPT-4 (the latest GPT version at the time of system development) regarding giving feedback on the images and generating post content (examples attached in the supplementary materials). We observed that GPT-4-0125-preview offers performance similar to GPT-4 in our prepared test cases. Therefore, GPT-4-0125-preview is a more economical choice for our use case. Moreover, both the frontend and backend have been deployed on an Alibaba Cloud server using Nginx and uWSGI, enabling remote user study and interviews with experienced beauty influencers. The frontend is in Chinese, therefore, the Edge translation plugin is employed to translate the website for the illustrations presented in this paper. Additionally, the decision to use Vue 2 was driven by the need to maintain compatibility with existing components (e.g., chatbot and RED Page Simulator) in the application, ensuring a smoother development process and minimizing disruptions.

4.3.1. Image selection

When users upload candidate cover images to the system, the images are fed into our model. The model then outputs the probability of each image being predicted as "will be clicked by viewers" and the top five important features with corresponding values. The probability is derived by computing the ratio of decision trees in the random forest that classify the sample as "will be clicked by viewers." It should be noted that the feature ranking process could be subject to biases. These biases may stem from the nature of the training data or the model's reliance on certain features that have been more prominent in past data, potentially affecting the accuracy of the rankings (Chandrashekar & Sahin, 2014). When users select an image for analysis, the probability of viewers' click intention on the image, as well as the values and rankings of the five important features, will be fed into the LLM prompt. The LLM prompt consists primarily of three parts. The first part serves as a contextual cue, indicating the identity of a beauty image analysis expert with unique insights into editing cover images for beauty notes in RED. The second part introduces our model, detailing the meanings of ten features, and the average values of positive and negative cases in our collected database. This part also clarifies the significance of feature importance rankings and includes some conclusions from user interviews obtained in the Section 3.3.3. The third part outlines the requirements for the output content, with three main points. First, the feature analysis should integrate the content of the second part. Second, the analysis and suggestions should demonstrate beauty-related knowledge, offering actionable and specific advice. Third, the output must fully adhere to the provided format. As seen in Figure 3 A2, *BeautyClicker* also offers the functionality of providing users with positive and negative cases for reference. When users choose to explore positive and negative cases of a feature such as "Face Area Percentage," our system selects "typical" cases from both groups based on their "Face Area Percentage" values in our database. For positive cases, we choose the images closest to the mean of all positive and negative cases, while for negative cases, we select the images nearest to the mean of all negative cases. This method enables users to focus on representative cases, thereby enhancing the user's understanding of how this feature manifests.

4.3.2. Text editor and chatbot

The content generation (title, content, Chatbot generation) is separately carried out by three LLM clients. *BeautyClicker*'s LLM clients are designed specifically for beauty product promotion, enabling them to generate content that is more relevant and aligned with the needs of beauty influencers compared to generic LLM-based assistants. By incorporating product features, selling points, and personal user input, along with considering cover image elements, the LLM client ensures that the generated titles and text are consistent with the cover images. This helps influencers craft posts that better match their intended message and audience expectations. Metadata does not preset a fixed content framework, but rather provides essential contextual information and situational cues for the generation process. For instance, when generating content for platforms like RED, metadata such as the post's topic and product features serve as contextual references for the model, rather than restricting the diversity of the content. As a result, metadata helps the model generate more flexible and varied outputs, ensuring that the content is both relevant to the user's input and aligned with the intended message. As seen in Figure 3 B2, beauty influencers can enter posts' metadata (the topic of the post, product features and selling points, and the user's feelings about the product). The prompt of three LLM clients all consists of three parts. The first part is all contextual cues, such as "You serve as a professional script generator for beauty posts in RED. Your task involves crafting sophisticated beauty product promotion posts based on user input." The second part delves into the significance of metadata and their diverse roles in content generation. For example, "{topic} represents the theme of the notes, which may be concise and precise. It could be the name of a product that a Little Red Book influencer recommends, such as lipstick, concealer, etc."

In particular, metadata encompasses more than just the three inputs entered by the user. It also includes the text identified on the cover image chosen by the user in the A1 panel. This approach could enhance the relevance among the cover image, title, and content of the post. However, it may decrease the diversity of the generated content as we prompt the LLM to take the metadata into considerations. Future work could systematically examine whether and how the inclusion of meta-data affect the diversity of generated content and study how the diversity of generated content affect beauty influencers' post outcome. The third section primarily offers format guidelines as references for the output content and provides both positive and negative cases of generated content, such as "Affordable Makeup! Prices as Low as 2 Yuan" as a postive case and "Heat-Resistant, Long-Lasting Makeup Hero" as a negative case. The detailed prompts used in *BeautyClicker* are attached in the supplemental material.

5. Evaluation

To explore the effectiveness and user experience of *BeautyClicker* for supporting the publication of beauty product promotion posts, we conduct two user studies. First, we conduct a within-subjects study that compares *BeautyClicker* to a baseline chatbot powered by GPT-4 with 32 participants who frequently browse beauty products in RED and have interests in editing a beauty product promotion post. Second, we conduct an expert interview with the same four beauty influencers in the design process in Section 4.1. Specifically, we examine how ClickMe affects the outcome of beauty product promotion posts, the process of editing these posts, and users' perceptions of ClickMe as a tool for assisting in the editing of beauty product promotion posts.

5.1. Within-subjects study

5.1.1. Participants

We determined the target sample size using G*Power to ensure sufficient statistical power for Wilcoxon Signed-Rank tests. Based on a two-tailed test with a normal parent distribution, a large effect size (0.8), an alpha level of 0.05, and a power of 0.95, the minimum recommended sample size was twenty-four. We ultimately recruited thirty-two females (P1–P32, age range 19–24, Mean = 21.32, SD = 0.95) from a local university via a post in a group chat and word of mouth. We recruit 32 females (P1-P32, age range 19-24, Mean = 21.32, SD = 0.95) from a local university via a post in a group chat and word of mouth. These participants are distinct from the twenty-eight annotators involved in the earlier annotation task. There are sixteen undergraduates, fourteen graduate students, one employee at an Internet company, and one employee at a state-owned enterprise. Thirty students major in engineering disciplines (e.g., Artificial Intelligence, Hydraulic Engineering, Computer Science, Computer Science and Technology, Geology), one in Philosophy, and one in the Korean Language. All participants are quite interested in having a trial on utilizing artificial intelligence tools or online resources for publishing product promotion posts (Mean =6.03, SD = 1.24; 1 - no interest at all, 7 - a large amount of interest). Nineteen participants have previously published beauty-related posts in RED, eleven participants have published other types of posts, and the remaining two participants have not posted content on social media platforms like RED. In particularly, all participants have experience browsing beauty-related posts on RED. Among them, eighteen participants do so almost every day, twelve browse two to six times per week, and two participants browse once a week. The detailed information of the 32 participants is placed in the Table 7 of Appendix B.

5.1.2. Baseline

The goal of the within-subjects study is to quantitatively evaluate how *BeautyClicker* can assist novice beauty influencers in optimizing their beauty product promotion posts. Compared to existing tools that support the drafts of posts, *BeautyClicker* is unique in the click intent prediction and analysis as well as the customized prompts for generating titles and content of beauty product promotion posts. Therefore, the **baseline tool** reserves the Image Upload feature (Figure 3 A1), the RED assistant (B3), the edit area of title and content in the Text Editor (B2), and the Preview panel (C1, C2) of *BeautyClicker* but does not have the buttons of "Upload for parsing" and "Sort by click probability," the Example panel (A2), and the Image Analysis panel (B1). The baseline functionality, such as the chatbot focused on generating RED-style promotional text, is comparable to existing market tools[6] that support influencer content creation. We designed it to mirror the general workflow and interface of common AI-assisted writing tools while excluding ClickMe's unique features, ensuring a meaningful and fair comparison. In either *BeautyClicker* or the baseline condition, participants were informed that they could search for anything on the official RED homepage if needed.

5.1.3. Tasks and procedure

Each participant has two post-editing tasks. The prompt for each task is:

> You are a novice beauty influencer. You receive an invitation from an advertising partner to promote the product d[Product Name in this task] via a post in RED. Your partners have provided you with a set of candidate cover images and a description of the product. You are required to select a cover image and write the title and content of the product promotion post. Now your partner wants to see your post in RED in 20 minutes.

We select two products that are essential for makeup application, *i.e.,* liquid foundation and loose powder. We balance the order of post-editing tasks and experienced tools, with five participants in each of the four assignments:

- *BeautyClicker* - liquid foundation, Baseline - loose powder
- Baseline - loose powder, *BeautyClicker* - liquid foundation
- *BeautyClicker* - loose powder, Baseline - liquid foundation
- Baseline - liquid foundation, *BeautyClicker* - loose powder

The whole procedure is illustrated in Figure 4. Before the day designated for post-editing tasks, we provided participants with a document introducing two promotion posts of other products in RED as examples, along with links to videos created by other influencers that illustrate the user experiences of the two products that participants are going to promote in our tasks. On the day of the post-editing tasks, each participant arrives at our lab or joins an online meeting. In each task, we first introduce the assigned system to the participants. We then assist participants in setting up their study environments on the provided computers or their own laptops, which included opening the webpages of the baseline / *BeautyClicker* interface and the RED homepage. Each task is allocated 20 min, and participants are informed that they can complete the task earlier or take a few extra minutes if needed. After each task, we ask participants to fill out a questionnaire regarding their experiences during the post-editing process and their perceptions of the system used. Subsequently, we conduct a 5-minute interview with them. Upon completion of both tasks, we conducted a final semi-structured interview focusing on their preferences regarding the systems used, perceptions of the generated content, opinions on the features

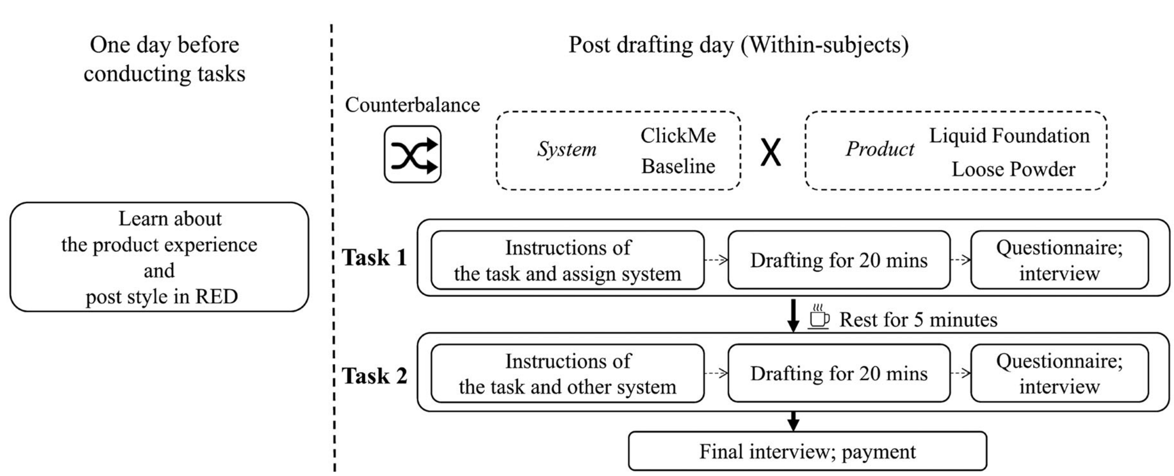

Figure 4. The procedure of the within-subjects study.

of *BeautyClicker*, and suggestions for its improvement. In total, each participant spend approximately 70 min in our experiment and received a compensation of 60 RMB.

5.1.4. Measurements

We employ a standard 7-point Likert scale (1/7 - strongly disagree/agree) to measure the quality of posts, experience in the process of editing product promotion posts, and perceptions towards the used tools.

5.1.4.1. Outcome of product promotion posts. To assess the quality of the outcomes, we invited two females, aged 21 and 22, who are familiar with the two product types used in our study. Both are avid users of the RED platform who browse beauty-related posts daily and are currently university students. They independently scored all outcome posts in a randomized order, providing ratings from the perspective of typical RED users, to ensure objective evaluation. For each product promotion post, the beauty influencer assesses it from three aspects adapted from previous work on persuasiveness (Mahmood & Huang, 2022; Pias et al., 2024; Thomas et al., 2019). These aspects are persuasiveness ("This post is persuasive"), relevance ("The content in this post is relevant to the product"), and trustworthiness ("The post is trustworthy"). The two annotators reached high levels of agreement on these three aspects, with ICCs (Koo & Li, 2016) of 0.803, 0.775, and 0.804, respectively. We average the scores of both annotators for each aspect.

5.1.4.2. Process of editing product promotion posts. Building on the NASATLX survey (Gatys et al., 2016), we present six questions designed to assess the workload experienced during the post-editing process. These questions cover various dimensions, including Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. In the interviews, we encourage participants to share how the tool has increased or decreased their workload. This allows us to gain a deeper understanding of the tool's effects on user experience and efficiency.

***5.1.4.3. Perception of* BeautyClicker.** Following the methodology outlined in Fan et al. (2024), we evaluate the usability of *BeautyClicker* and the baseline tool using the ten questions from the System Usability Scale (SUS) (Brooke, 2013), which are incorporated into our questionnaire. We divide SUS into three aspects: Effectiveness & Learnability, Use Efficiency, and Satisfaction. Besides, we adapt five questions from Jian's Trust Scale (Jian et al., 2000) and investigate how users trust the used tool by inquiring about the tool's vigilance, potential negative impact on teaching, their trust in the system's ethical standards, commitment to users, and the reliability of outputs. In the interviews, we encourage them to share their experiences with each tool and preferences for each component (*e.g.,* different preset activities, the image analysis panel in *BeautyClicker*).

5.2. Expert interview

Apart from the within-subjects study that evaluates *BeautyClicker*'s effectiveness in assisting novices, we further evaluate *BeautyClicker* with the four beauty influencers (I1-4) who participate in our design process (Section 4.1). We conduct a 60-minute interview with each influencer. After reviewing the background of *BeautyClicker*, we give a 10-minute tutorial on how to use it. Then, participants spend 25 min to complete two tasks and are asked to think aloud in this process. The two tasks are selecting a cover image from candidate images and composing the title and content for promoting the liquid foundation product used in the within-subject study. Finally, we conduct a 25-minute semi-structured interview with guiding questions focused on three areas: evaluating the quality of cover images and content generated by *BeautyClicker*, assessing its impact on editing efficiency and workload, and gathering user perceptions and suggestions for improvement. Furthermore, we invite one of the influencers (I1) to use our tool in her real work context. Specifically, she is tasked with selecting a cover image and composing a title and caption for a post she is preparing. During this process, I1 is asked to think aloud while using the tool, providing us with valuable insights into her experience and the effectiveness of *BeautyClicker* in a real-world scenario. This helps us further assess the tool's usability and performance in actual content creation.

6. Analyses and results

In this section, we present the quantitative and qualitative results for each research question. As for the items measured about the outcome posts, the process of editing the product Promotion posts, and perceptions of *BeautyClicker* in the within-subjects study, We conducted a set of mixed-ANOVA tests to examine the effects of tool usage (Tool), tool usage order (Order), and task presentation (Task) on 19 outcome measures (Table 6 in Appendix A). A total of 133 comparisons are made across the following factors: Tool, Order, Task, Tool * Order, Tool * Task, Tool * Order * Task, and Order * Task. Tool's main effect was significant for most variables, which suggests the choice of tool significantly impacts user performance and perception. Among other effects, only 7 showed statistically significant effects ($p < 0.05$), including Temporal demand (*Task*: $p = 0.012$), Own Performance($Order * Task : p = 0.015$), Effort(Tool*Order:$p = 0.035$), Efficiency in use(Order: $p = 0.038$), Trust in Ethical Standards ($Tool * Task : p = 0.032$), Potential Negative Impact ($Tool * Order : p = 0.039$), and Vigilance (*Task*: $p = 0.042$). These effects are reasonable in a within-subjects study, as participants may become more familiar with the task and tools during the process. However, given the low number and scattered distribution of significant effects, we believe that order doesn't meaningfully bias the overall results. Thereafter, we only focus on the tool as the independent variable in the analyses and employ the Wilcoxon signed-rank test (Woolson, 2008), a non-parametric test, to compare the differences between the ClickMe and baseline tools. Given our within-subjects design and the use of Likert-scale responses, this test provides a robust alternative to the paired t-test without assuming normality, and is widely used in prior HCI research (Kang et al., 2021; Peng et al., 2024). For the qualitative data in the within-subjects study and expert interviews, two authors transcribe the audio into text scripts and conduct a thematic analysis on these scripts. They first familiarize themselves by reviewing all the text scripts independently. After several rounds of coding with comparison and discussion, they finalize the codes of all the interview data based on three aspects: the outcome of beauty product promotion posts, the process of editing such posts, and users' perceptions of *BeautyClicker*. We count the occurrences of codes and incorporate these qualitative findings in the following presentation of our results.

6.1. The analyses and results of within-subjects study

6.1.1. Outcome product promotion posts

Table 5 summarizes the statistical results in the within-subjects study. Overall, the product promotion posts created using *BeautyClicker* (Mean = 15.17, SD = 2.24) are rated significantly better than those developed with the baseline system (Mean = 11.86, SD = 2.60); $Z = -4.49$, $p < 0.001$. Specifically, the posts created with *BeautyClicker* are significantly more persuasive (*BeautyClicker*: Mean = 4.45, SD = 1.23; Baseline: Mean = 2.84, SD = 1.14; $Z = -4.61$, $p < 0.001$), more relevant to the product description (*BeautyClicker*: Mean = 6.08, SD = 0.46; Baseline: Mean = 5.20, SD = 0.79; $Z = -4.19$, $p < 0.001$), and more trustworthy (*BeautyClicker*: Mean = 4.64, SD = 1.12; Baseline: Mean = 3.81, SD = 1.29; $Z = -2.54$, $p = 0.011$). Specifically, regarding the selected cover images in the outcome posts, 27 participants with *BeautyClicker* choose the image that is also recommended by our experienced beauty influencers, while only 8 participants in the baseline conditions are able to do so. The experienced influencers' choices on the best cover images for the two products also have the highest probability predicted by our click intent prediction model. This illustrates that our model can assist novices in selecting better cover images. P5 indicated, *"The probability of being clicked and the explanation regarding the feature of Face Area Percentage helps me make the best choice."* Besides, four participants (P1, P7, P11, P15) mention that compared to the baseline tool, the post's textual content generated by our system is less flamboyant and more authentic.

Table 5. Statistical results about *BeautyClicker* and the baseline interface.

Research question	Item	*BeautyClicker* Mean (SD)	Baseline Mean (SD)	Statistics Z	p	Sig.
Outcome of Product Promotion Posts	Sum	15.17 (0.40)	11.86 (0.46)	−4.491	< 0.001	***
	-Persuasiveness	4.60 (0.22)	2.84 (0.20)	−4.612	< 0.001	***
	-Relevance	6.13 (0.08)	5.20 (0.14)	−4.187	< 0.001	***
	-Trustworthiness	4.65 (0.20)	3.81 (0.23)	−2.544	0.011	*
Process of Editing Product Promotion Posts	Mental Demand	1.59 (0.10)	3.78 (0.35)	−4.268	< 0.001	***
	Physical Demand	1.34 (0.09)	3.00 (0.35)	−3.843	< 0.001	***
	Temporal Demand	1.50 (0.16)	3.75 (0.35)	−4.351	< 0.001	***
	Own Performance	6.28 (0.14)	4.09 (0.26)	−4.167	< 0.001	***
	Effort	2.38 (0.21)	4.56 (0.33)	−3.764	0.008	**
	Frustration	1.38 (0.11)	2.56 (0.30)	−3.469	0.005	**
Perceptions of *BeautyClicker*	System Usability Scale	84.90 (1.74)	62.09 (4.07)	−4.240	< 0.001	***
	-Effectiveness & Learnability	84.72 (2.11)	63.54 (4.11)	−4.025	< 0.001	***
	-Efficiency in Use	86.32 (2.09)	65.10 (4.86)	−3.458	< 0.001	***
	-Satisfaction	83.15 (2.24)	56.59 (4.09)	−4.404	< 0.001	***
	Vigilance	1.67 (0.09)	2.94 (0.32)	−3.541	< 0.001	***
	Potential Negative Impact	1.28 (0.09)	2.22 (0.30)	−2.953	0.003	**
	Trust in Ethical Standards	6.28 (0.26)	5.72 (0.33)	−2.098	0.036	*
	Commitment to Users	6.28 (0.12)	4.50 (0.29)	−3.867	< 0.001	***
	Reliability of Outputs	6.25 (0.13)	4.50 (0.25)	−4.196	< 0.001	***

All items are measured using a standard 7-point likert scale (1 - strongly disagree; 7 - strongly agree). *Note:* −: $p > 0.1$, +: $0.05 < p < 0.10$, *: $p < 0.05$, **: $p < 0.001$,***: $p < 0.01$; wilcoxon signed-rank test; within-subjects; $N = 32$.

6.1.2. Process of editing product promotion posts

As shown in Table 5, compared with the baseline condition, participants with *BeautyClicker* perceived significantly less mental demand (*BeautyClicker*: Mean = 1.59, SD = 0.10; Baseline: Mean = 3.78, SD = 0.35; Z = −4.268, $p < 0.001$), physical demand (*BeautyClicker*: Mean = 1.34, SD = 0.09; Baseline: Mean = 3.00, SD = 0.35), temporal demand (*BeautyClicker*: Mean = 1.50, SD = 0.16; Baseline: Mean = 3.75, SD = 0.35), and frustration (*BeautyClicker*: Mean = 1.38, SD = 0.11; Baseline: Mean = 2.56, SD = 0.30; Z = −3.469, $p = 0.005$) in the process of editing product promotion posts. When they are with *BeautyClicker*, participants also experience less effort (*BeautyClicker*: *Mean* = 2.38, *SD* = 0.21; Baseline: Mean = 4.56, SD = 0.33, $p = 0.008$) and feel that they have better performance (*BeautyClicker*: Mean = 6.28, SD = 0.14; Baseline: Mean = 4.09, SD = 0.26, $p < 0.001$). Twenty-eight participants express that providing click-through predictions on the cover images, together with the analyses and suggestions on the image features, can alleviate the burden of thinking when selecting a cover. For instance, six participants (P1, P8, P12, P25, P28, P29) mention that the high score of viewers' click intent on the selected cover image reduced their hesitations during the image selection stage. P7 states, "*I made the decision faster with the click intention predictor, especially when I was stuck between multiple* images. I can *easily rule out the images with the lowest probabilities.*" However, one participant (P11) has a contrary view. Although certain images do not have high scores from the predictor, P11 feels that the facial expressions and textual content within these images could evoke her emotional resonance. This mismatch between the prediction and P11's feelings make her more hesitant in choosing images. Besides, P4, P8, and P28 also mention that, as novice beauty influencers, the cover image modification suggestions provided are somewhat abstract and difficult for them to understand. Apart from the image analysis panel, twenty-seven participants also favor the text editor with customized support that alleviates their burdens in editing creative content in the beauty product promotion posts. For example, P6 saies, "*When it came to the stage of crafting the post's content, I found myself at a loss for words. The generated content (in BeautyClicker) served as a wellspring of inspiration and guidance for me.*"

6.1.3. Perceptions of BeautyClicker

Participants' perceptions of the *BeautyClicker* system are generally positive. Specifically, participants rate *BeautyClicker* (Mean = 84.90, SD = 1.74) significantly higher than the baseline tool (Mean = 62.09, SD = 4.07) in terms of usability; Z = −4.240, $p < 0.001$. *BeautyClicker* also outperform the baseline tool in the sub-aspects of the usability scale, including learnability, efficiency in use, and satisfaction ($p < 0.001$, Table 5). As for the trust to the tool, compared to the baseline tool, participants with

BeautyClicker generally have significantly fewer concerns regarding its vigilance (*BeautyClicker*: Mean = 1.67, SD = 0.09; baseline: Mean = 2.94, SD = 0.32; $Z = -3.541$, $p < 0.001$) and potential negative impact on editing post (*BeautyClicker*: Mean = 1.28, SD = 0.09; baseline: Mean = 2.22, SD = 0.30; $Z = -2.953$, $p = 0.003$). Participants also have a significantly higher level of trust on *BeautyClicker*'s ethical standards (*BeautyClicker*: Mean = 6.28, SD = 0.26; baseline: Mean = 5.72, SD = 0.33; $Z = -2.098$, $p = 0.036$), commitment to user (*BeautyClicker*: Mean = 6.28, SD = 0.12; baseline: Mean = 4.50, SD = 0.29; $Z = -3.867$, $p < 0.001$), and reliability of outputs (*BeautyClicker*: Mean = 6.25, SD = 0.13; baseline: Mean = 4.50, SD = 0.25; $Z = -4.196$, $p < 0.001$), suggesting that *BeautyClicker* is perceived as a reliable and user-friendly platform. Moreover, compared to the baseline ($N = 9$), a greater number of individuals ($N = 31$) report in the interview after each task that they feel confident that their outcome posts would have a high click-through rate. Six participants (P1, P2, P4, P5, P23, P29) mention that the cover image selected by *BeautyClicker* infuses them with confidence to publish their posts. However, P12 expresses that although the model tends to prefer visually appealing images for cover selection, the actual user's perception and click behavior exhibit a certain level of uncertainty and randomness. P32 states: "The suggestions related to the features derived from the cover analysis may appear somewhat rigid, as the performance of these features may need to be adjusted and determined based on the specific characteristics of different products." At the end of the interview, all participants say that *BeautyClicker* helped them learn the knowledge relevant to beauty product promotion posts. For instance, P10 mentions she has acquired knowledge about what key features should be considered when creating appealing cover images.

The four experts agree that click intent prediction is their favorite feature in *BeautyClicker*. In particular, I1 expresses her intention to utilize this feature in the future, acknowledging that the feature of cover selection has alleviated her burden of thinking in her trial. I3 also recognizes the substantial assistance provided by the image analysis feature. She states, "*This feature has transformed a rather abstract image into quantifiable data. It enhances rational thinking while observing it, potentially leading to more accurate outcomes.*" Four experts have also given some suggestions regarding *BeautyClicker*. I1 claims that the preview panel (C2) can showcase all the candidate cover images, which could give a visual comparison. I2 expresses a desire to generate multiple content or titles simultaneously for her selection. I3 mentions a wish for our Image Analysis panel to offer visual representations of images improved based on the suggestions provided. I4 mentions the desire to categorize the positive and negative cases based on the types of beauty products.

6.2. The analyses and results of expert interview

6.2.1. The interview of four beauty influencers

All four beauty influencers acknowledge that the model's selection closely aligns with the cover image they come up with. For instance, I2 state, "*I would pick this one too. It strikes a nice balance with a bit of a face but not too overwhelming, and it showcases the product well. The colors and brightness are just right, making it look perfect.*" All experts also acknowledge the effectiveness of our Text Editor Panel. In particular, I1 expresses, "*The chatbot's text seemed ignorant about makeup, but the text from your system reads like it is written by someone who knows about beauty terms.*"

Furthermore, the four experts also reach a consensus on efficient cover image selection through click intent prediction and image analysis. Especially, I2 reports, "*Regardless of whether this answer provided by the model is my final choice, it gives me a shortcut for choosing a cover image for my posts in RED.*" Besides, I3 states, "*I did not realize that some image features like Orange Element Proportion could affect the click intent of my posts. I will keep that in mind when making or choosing covers next time.*" Regarding the Text Editor panel, I1, I3, and I4 mention that the generated content provides a framework for their content creation, allowing them to add more specific details.

Finally, the four experts agree that click intent prediction is their favorite feature in *BeautyClicker*. In particular, I1 expresses her intention to utilize this feature in the future, acknowledging that the feature of cover selection has alleviated her burden of thinking in her trial. I3 also recognizes the substantial assistance provided by the image analysis feature. She states, "*This feature has transformed a rather abstract image into quantifiable data. It enhances rational thinking while observing it, potentially leading* *to more accurate outcomes.*" Four experts have also given some suggestions regarding *BeautyClicker*. I1 claims that the preview panel (C2) can showcase all the candidate cover images, which could give a visual comparison. I2 expresses a desire to generate multiple content or titles simultaneously for her selection. I3 mentions a wish for our Image Analysis panel to offer visual representations of images improved based on the suggestions provided. I4 mentions the desire to categorize the positive and negative cases based on the types of beauty products.

6.2.2. Exploring the use of BeautyClicker in real-world content creation

In this section, we present the insights gained from inviting one of the influencers (I1) to use *BeautyClicker* in her real-world content creation process. The process of the beauty influencer (I1) utilizing our tool for selecting a cover image and creating content is illustrated in Table 5. I1 reports that the cover analysis panel's feature of providing click data helped alleviate some of her hesitation when selecting a cover image. However, she expresses concerns about the accuracy of the model's ratings. Despite this, she acknowledges the effectiveness of the system's feature analysis, particularly regarding color-related features. She notes that the system helped her recognize aspects of the image's color scheme that she might have overlooked and provided valuable suggestions for editing the image. I1 also highlights the practical significance of the cover feature analysis for novice influencers. She points out that a key difference between novice beauty influencers and their more experienced counterparts lies in image shooting and editing. The feature analysis, along with the positive and negative examples provided, enables novice influencers to quickly identify gaps between their work and that of experienced influencers, thus helping them bridge these gaps.

When discussing the title and caption generation feature, I1 mentions that the generated content still fell short compared to what professional beauty influencers typically produce, particularly in keeping up with current trends and slang. Nonetheless, she acknowledges that the content generation feature could serve as a useful source of inspiration for beginners, especially for those who are just starting out in the beauty influencer space. Finally, I1 emphasizes that, in her view, AI-generated content, particularly in the realm of professional beauty writing, could never fully surpass human creativity and expertise (Figure 5).

7. Discussion

7.1. Reflection on computational modeling for viewers' click intent

In our work, we propose a data-driven approach to predict viewers' click intent on beauty product promotion posts in RED, with the best model achieving the precision of 0.7554 and F1 score of 0.5894 using ten features of the post's cover image as input. Via the interviews with both RED users and beauty influencers, we make sense of the impact of these ten image features, including the color-based (*e.g.,* orange element proportion), texture-based (*e.g.,* contrast), and content-based (*e.g.,* face area percentage), on predicting viewers' click intent. These empirical findings support the results of previous studies on user perceptions of images (Bakhshi et al., 2014; Elliot & Maier, 2007; Fink et al., 2001). For instance, Fink et al. (2001) demonstrate that the texture features of women's skin in images influence the perceived beauty of the female face in those images. Furthermore, Bakhshi et al. (2014) find that images with faces can attract more likes on Instagram. Our model fills the blank of modeling viewers' click intent on product promotion posts.

However, our model also has examples of failure classification. Figure 6 presents four examples with the true labels provided by crowd workers and the predicted labels generated by our model. Analysis of these misclassified posts suggests that the failure to accurately identify posts may be attributed to two main factors. First, the model may not adequately comprehend the combinational relationships between features. For instance, when a post has low values for both the proportion of the face and the distribution of orange regions (*e.g.,* Figure 6(b)), the model struggles to integrate other features beyond these two, leading to classification failure. This aspect can refer to the research by Li et al. (2024), which considers the relationships between multiple features. In the future, it may be beneficial to consider multi-feature modeling, such as combining content features with color features.

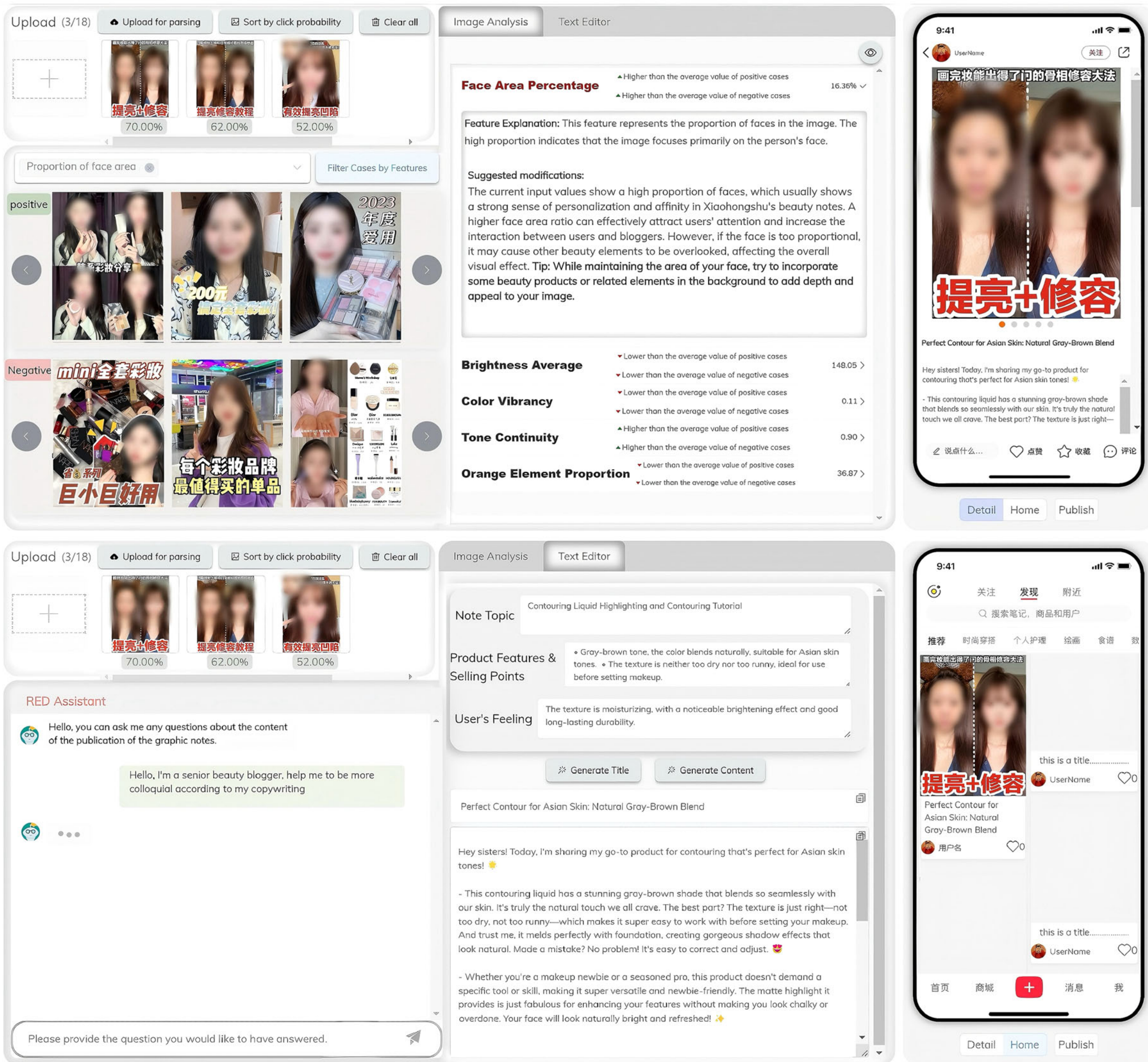

Figure 5. The beauty influencer (I1) utilizes our tool for selecting cover image and creating content prior to publishing post.

Second, there are potentially influential features affecting click intention on posts that have not yet been unearthed. For example, in Figure 6(c) and (d), despite excelling in features like face area percentage, they are still not chosen by the annotators. For example, during interviews with six RED users in Section 3.3.3, some mention that although the cover images of some posts are attractive, based on their experience, these posts may be suspected of being product promotion. This suspicion can affect the intention to click. Besides, they also raise several such features, such as viewers' personal needs like "budget-friendly" and background information like "yellow skin tone." Previous research has offered insights to incorporate such user factors into modeling (Peng et al., 2021; Zhang & Pennacchiotti, 2013). For instance, Zhang and Pennacchiotti (2013) incorporated features about users' social media activities when constructing models to predict purchasing behaviors. Accordingly, when available, the models should incorporate features about viewers as input.

7.2. Reflection on BeautyClicker

In our work, we also develop *BeautyClicker* that explores the value of our model in helping beauty influencers select and analyze cover images for their product promotion posts. Our within-subjects study and expert interviews both confirm that our model can reduce users' hesitation in selecting a cover image while making a better choice at the same time. Our work extends previous work on

(a) True Label:"1" Predicted Label:"0"
(b) True Label:"1" Predicted Label:"0"
(c) True Label:"0" Predicted Label:"1"
(d) True Label:"0" Predicted Label:"1"

Figure 6. Examples of model classification failures. (Note: "true label" is the labels provided by crowd workers and "predicted label" is the predicted labels generated by our model).

supporting the textual drafts of social media posts (Aldous et al., 2019; 2024; De & Lu, 2024; Peng et al., 2020b) by introducing computational models to facilitate image selection. Our computational approach and *BeautyClicker* can serve as starting points for assisting the drafts of posts that promote other types of products or ideas on social media, *e.g.*, electric products and clothes (Kumar & Alok, 2020; Zafar et al., 2021). Future work on image generation for beauty products could also leverage the click intent prediction from our model as one metric to evaluate the generated images. However, our user study findings also highlight the opportunities of *BeautyClicker* from the perspective of the viewers and the poster, which we summarize as design considerations for future interactive systems that support drafts of beauty product promotion posts.

7.2.1. User profile

The first design consideration is that, before uploading an image, the system should provide an option allowing the influencer to select the target audience for the post. This corresponds to Section 7.1 of incorporating user factors into the modeling process. The implementation of this feature can be guided by previous approaches (Ma et al., 2021; Qian et al., 2021). Before uploading an image, the influencer can select the target audience for the post such as a student demographic through the system options. The influencer can input the target user profile, which will be integrated into the modeling process to guide LLM in offering relevant cover image improvement suggestions. To enhance this feature, the system could be designed to dynamically adjust its suggestions based on the influencer's continuous interactions with the system (Reviglio della Venaria, 2020). By analyzing the performance of previous posts targeted at specific audiences, the system can refine its recommendations over time. For example, it could consider which features, such as image color schemes or the tone of the text, led to higher engagement within the selected target group. Additionally, the system can predict and return the click intentions of both the target and non-target user groups.

7.2.2. Aesthetic fatigue

During the interviews with six RED users in Section 3.3.3, they also raise the issue of aesthetic fatigue regarding social media post covers. The cover images on the recommendation posts are often very exquisite and beautiful, which might lead to a sense of over-saturation for some users who scroll through social media daily. The users point out that the sameness of cover images is one of the factors leading to aesthetic fatigue. This trend towards homogenization has been confirmed in influencers' interviews in Section 4.1.2, as the influencer mentions imitating the cover designs of popular posts. This approach is likely based on the pursuit of success, with the belief that replicating the appearance of popular content can help their own posts garner more attention (Gräve & Greff, 2018). However, when many influencers adopt similar strategies, social media becomes saturated with stylistically similar cover images. This not only reduces the diversity of content but can also lead to user fatigue due to the repetitiveness of the content (Zhao et al., 2024). Such aesthetic fatigue may prompt users to seek out content that is more personalized and original, as it may appear fresher and more authentic both visually and emotionally. Therefore, in the future design of systems, modeling for cover images should consider factors from multiple dimensions, such as the fun and novelty of the image, to make predictions and recommendations for cover images more comprehensive and diverse.

7.2.3. Visual enhancements and cover image template

For content creators, there is still room for improvement in our system. First, a post assistant like *BeautyClicker* should support automatic visual enhancements on the candidate cover images. While our image analysis panel contains LLM-generated suggestions on improving certain features in the selected cover image, such suggestions are rather abstract, as indicated in Section 6.1.2. To provide more concrete suggestions, future tools could consider employing automatic visual enhancement techniques, such as Image Style Transfer (Gatys et al., 2016) and Stable Diffusion (Yang et al., 2024). In addition, the interviews with beauty influencers in Section 4.1.2 indicate that for influencers with many years of experience, their cover images tend to develop a specific style. *BeautyClicker* could add a new feature that is capable of learning and analyzing the past cover designs of the influencer, thereby creating a specific template. This type of design template would not only reflect the unique style of the influencer but also enhance the recognizability of their posts. In detail, when an influencer uploads a new, unedited cover photo and enters the cover text, the system would be able to automatically generate an edited cover photo based on the influencer's historical cover templates. This would involve placing the text in the style that the influencer usually prefers and adjusting the new photo to match the color schemes and composition patterns of the historical covers. Such a feature would not only save the influencer precious time but also help them maintain a consistent brand image in their content creation, which is essential for attracting viewers and increasing engagement. To further improve this, the system could incorporate an adaptive learning model (Gligorea et al., 2023) where the generated templates are not only based on historical designs but also on the influencer's current preferences. This would allow the system to stay updated with evolving design trends or changing branding strategies. Additionally, feedback mechanisms could be introduced, allowing influencers to provide ratings on how well the system-generated templates align with their vision. This would enable the system to fine-tune its output accordingly. This function can be implemented by using existing image generation technology. These technologies, such as generating a confrontation network (GANS) (Li et al., 2021), can analyze the models in existing data and generate new content with similar styles and aesthetics.

7.3. Generalizability of our work

Although our model focuses on beauty product promotion posts in RED, our approach can also be applied to other product promotion posts such as organic products (Pechová et al., 2015). The model can be modified based on the specific characteristics of different products, allowing us to customize *BeautyClicker* accordingly. Besides, some elements of the model can also be applied to other categories of posts on different social media platforms. For instance, color-related features can be used to predict click intention for travel diary posts on Instagram, just as Yu et al. (2020) research found that the hue and brightness of pictures can affect the popularity of travel-related posts on Instagram, including user likes and commenting behavior when it comes to posts like travel journals, the facial features on the cover image have a minimal impact on user click intent. Consequently, when modeling such posts, it may be deemed unnecessary to take into account the processing of such features. Overall, our work can find wide-ranging applications in predicting users' click intent in social media post images, assisting influencers in enhancing their content creation. Nevertheless, researchers should appropriately modify the set of relevant features by either removing or adding them as necessary.

7.4. Ethical concerns

7.4.1. Clickbait

Clickbait is a marketing strategy designed to capture attention and entice users to click. Clickbait is prevalent on social media platforms. At times, it is defined as intentionally misleading audiences about the content of an advertisement or news, often through headlines, images, or titles (Bazaco et al., 2019). The issue with clickbait lies in its use of bait-and-switch tactics. It immediately grabs attention but fails to deliver upon user clicks, which will lead to user frustration or even dissatisfaction. Social media platforms should proactively take measures to identify deceptive clickbait practices. This may involve enhancing algorithms to detect misleading or exaggerated content (Zuhroh & Rakhmawati, 2020), implementing corresponding policies for content creators, and raising user awareness about clickbait. However, not all clickbait is inherently bad. When used appropriately, it can serve as a potent supplement for influencer promotion. For influencers, it is essential to exert more diligent efforts in content creation, captivating audiences without compromising authenticity. This approach helps build trust and credibility, effectively leveraging the principle of attraction to enhance content appeal while maintaining quality.

7.4.2. Credibility of generative-AI

generative-AI has been widely used by influencers for social media content creation (Lyu et al., 2024), including the generation of text, images, and videos. Our system uses generative AI to generate suggestions for the influencer's cover image. When using generative-AI to generate suggestions for modifying cover images, the issue of credibility presents an ethical challenge (Huschens et al., 2023). Generative-AI generates recommendations by analyzing large datasets, but its outputs are not always reliable or accurate. The "black-box" nature of generative-AI makes it difficult for users to fully understand its decision-making process, thereby reducing its credibility. AI may also harbor biases (Fan et al., 2024), as its training data may include unjust styles or cultural prejudices, leading to recommendations that lack diversity and inclusivity. Therefore, in practical applications, it is crucial to strike a balance between AI's assistance and human judgment to ensure that the generated recommendations are both appropriate and credible. Besides, our system also involves the use of generative-AI to create product promotional posts. The AI-generated text may exaggerate the effects of the products, leading to unrealistic consumer expectations. Such misleading information could harm consumer interests. Influencers should carefully review AI-generated text to ensure its accuracy and truthfulness. Additionally, technological tools could be developed to detect potential biases or misleading information in AI-generated content.

7.4.3. Ethical implications

Due to the characteristics of the data types and annotator categories, our model may involve certain ethical risks. Therefore, it is essential to carefully consider and appropriately address these potential issues when applying the model. First and foremost, system developers must pay particular attention to potential biases within the model. For instance, the training data should ideally encompass a diverse set of images, with special attention given to ensuring the representation of individuals from different racial backgrounds. If certain groups are underrepresented in the image samples, the model may struggle to make accurate predictions when faced with less common cases. It is especially critical for underrepresented groups. In fact, many current algorithms tend to perform poorly when identifying images of these groups (Bennett et al., 2021). Beyond racial biases, gender bias is another significant ethical concern. If the training data and model design are heavily influenced by female-centric beauty standards or predominantly focus on female beauty content, the model may fail to adequately recognize and predict behaviors or preferences for male influencers and viewers.We develop our model and evalaute *BeautyClicker* with female participants as the beauty industry has historically been more female-oriented. However, this skew in data can lead to models that reinforce gender stereotypes, such as those that prioritize women's beauty products while neglecting male-centered beauty content. To mitigate this issue, it is crucial that future models incorporate a diverse set of gender perspectives in their training data and annotator pool. By accounting for both racial and gender diversity, we can mitigate the risk of the model reinforcing existing stereotypes or excluding certain groups.

7.5. Limitations and future work

Our work also has several limitations that call for future work. First, we use two common types of beauty products to explore *BeautyClicker*'s effectiveness in assisting novices in drafting product promotion posts. Further research is necessary to validate the findings of our within-subjects study across a broader range of beauty products. Simultaneously, the cultural context of RED, predominantly reflecting Chinese beauty standards, poses a particular limitation. Future studies should aim to include a more diverse cultural sample to evaluate the applicability of our model across varied cultural contexts, ensuring that the findings are broadly relevant and not solely applicable to the specific user base of RED.

Second, our research primarily focuses on female influencers and viewers. Although young females dominate the RED platform in beauty, perspectives from male beauty content creators and male viewers should not be neglected. As the beauty industry becomes increasingly inclusive (Hudders & De Jans, 2022; Sharma et al., 2024), future research should incorporate data on male beauty products and behaviors, and include male viewers, influencers, and annotators. For example, adding male beauty products and content from male influencers could provide a more balanced dataset that reflects the diversity of user preferences. This would allow the model to account for gender-specific visual preferences and promotional strategies, helping to tailor click intent prediction more effectively. Additionally, integrating male-centered beauty content into the model can improve its applicability to a broader audience and ensure its relevance across different gender groups.

Third, we base on a relatively small dataset concerning beauty product promotion in RED to model viewers' click intent, which may limit the applicability of our findings to multi-platform creative practices. For future work, we suggest expanding the data sources used for modeling viewer's click intentions, while also incorporating viewer-specific factors to achieve more accurate predictive outcomes. At the same time, the annotators in our study come from a relatively homogeneous background. Therefore, in future work, it would be beneficial to include annotators from more diverse backgrounds. This would help reduce individual biases and improve the overall quality of the annotations. Moreover, our qualitative insights are based on interviews with six RED users and four beauty influencers, both of which are limited in sample size. While their feedback provides valuable perspectives on model interpretability and practical relevance, the small number of participants restricts the generalizability of our findings. Additionally, as participants were asked to evaluate features pre-selected by the model, their responses may have been subject to confirmation bias. Future research should consider engaging a more diverse and larger participant pool and adopting blind or open-ended evaluation methods to mitigate such bias and strengthen the validity of the results. Furthermore, although our annotation approach was based on previous research (Guo et al., 2022b; Peng et al., 2020b), our annotation approach don't account for timing factors, such as when users view posts or how the ranking of posts on the platform may influence the likelihood of clicks. Given that platform algorithms often personalize recommendations based on user behavior, these factors could play a important role in click intent. Therefore, future work could explore collaboration with platforms to incorporate timing and personalized recommendation factors, thereby improving the accuracy of click intent predictions.

Meanwhile, the cover image modification suggestions provided by our system have not yet been validated through field study. Future research could further explore this aspect to evaluate the effectiveness of these suggestions in improving the first impressions of cover images in RED. Additionally, a significant limitation of our study is the absence of direct use of post views metrics for modeling click intent, primarily due to platform restrictions that limit access to these metrics. Incorporating such metrics could provide a more robust and directly relevant measure of engagement, enhancing the predictive accuracy and applicability of our model. Future work should aim to integrate CTR data, where accessible, to refine and validate the effectiveness of predictive models like *BeautyClicker* in real-world scenarios. At the same time, a limitation of our study is that we were unable to measure whether our tool actually improved clicks, as we don't have permission from brand owners to allow users to post their content in the RED and assess real-world performance. As a result, our current study focuses on evaluating the tool's usability and perceived value, but the direct impact on click-through rates remains untested. Future work could involve collaborating with brand partners to conduct a more comprehensive experiment. In this experiment, one group of new users would post without using our tool, while another group would use the tool to create their posts. By comparing the click-through rates of posts from both groups, we could evaluate whether our tool leads to an actual improvement in clicks. This would provide valuable insights into the practical effectiveness of the tool in real-world content creation. Lastly, we concentrate exclusively on the cover of the product promotion post. However, promotional posts typically include not only cover images and titles but also additional images and video formats (*e.g.*, YouTube, TikTok). This multi-modal content could also influence user engagement and the effectiveness of promotional posts (Shahbaznezhad et al., 2021).

8. Conclusion

In this paper, we propose a data-driven approach to predict viewers' intentions to click on beauty product promotion posts in RED based on the features of their cover images. Drawing from interviews with four beauty influencers, we develop *BeautyClicker*, a web application powered by our model and the large language model (LLM) GPT-4 to support the drafts of beauty product promotion posts. The within-subjects study involving twenty participants demonstrates that posts generated with *BeautyClicker* were perceived as significantly more persuasive, trustworthy, and relevant to the products. Additionally, participants report feeling more confident and less hesitant about their choices of cover images. Follow-up interviews with four experienced beauty influencers involved in our design process further affirm the value of our model in supporting cover image selection. We reflect on the insights gained from our computational model and user studies in modeling viewers' click intent and supporting the creation of social media posts.

Notes

1. https://www.xiaohongshu.com.
2. https://www.qian-gua.com/blog/detail/2898.html.
3. https://github.com/cv-cat/Spider_XHS.
4. https://www.wjx.cn.
5. https://ai.baidu.com/ai-doc/.
6. https://lobechat.com/discover/assistant/xiaohongshu.

Author contribution statement

Xueer Lin: Conceptualization, Methodology, System develop ment, User study, Data analysis, Writing– original draft, Writing review & editing. Xia Chen: System development, User study, Data analysis, Writing– original draft. Zhenhui Peng: Writing– review & editing, Data analysis, Supervision, Funding acquisition.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the Young Scientists Fund of the National Natural Science Foundation of China (NSFC) with Grant No. 62202509, NSFC Grant No. U22B2060, and the General Projects Fund of the Natural Science Foundation of Guangdong Province in China with Grant No. 2024A1515012226.

ORCID

Xueer Lin http://orcid.org/0009-0006-0404-9372
Xia Chen http://orcid.org/0009-0003-6300-8966
Zhenhui Peng http://orcid.org/0000-0002-5700-3136

References

Abidin, C., & Ots, M. (2016). Influencers tell all? Unravelling authenticity and credibility in a brand scandal. In M. Edström, A. T. Kenyon, & E.-M. Svensson (Eds.), *Blurring the lines: Market-driven and democracy-driven freedom of expression* (pp. 153–161). Nordicom.

Aldous, K. K., An, J., & Jansen, B. J. (2019). *The challenges of creating engaging content: Results from a focus group study of a popular News Media Organization* [Paper presentation]. Extended Abstracts of the 2019 CHI conference on human factors in computing systems (pp. 1–6), (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3290607.3312810

Aldous, K., Salminen, J., Farooq, A., Jung, S.-G., & Jansen, B. (2024). *Using ChatGPT in content marketing: Enhancing users' social media engagement in cross-platform content creation through Generative AI* [Paper presentation]. Proceedings of the 35th ACM conference on hypertext and social media (pp. 376–383), New York, NY, USA. (Poznan, Poland) *(HT '24)*. Association for Computing Machinery. https://doi.org/10.1145/3648188.3675142

Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, *9*(5), 272.

Azimi, J., Zhang, R., Zhou, Y., Navalpakkam, V., Mao, J., & Fern, X. (2012). *Visual appearance of display ads and its effect on click through rate* [Paper presentation]. Proceedings of the 21st ACM international conference on information and knowledge management (pp. 495–504), New York, NY, USA. (Maui, Hawaii, USA) *(CIKM '12)*. Association for Computing Machinery. https://doi.org/10.1145/2396761.2396826

Bahtar, A. Z., & Muda, M. (2016). The impact of user–generated content (UGC) on product reviews towards online purchasing—A conceptual framework. *Procedia Economics and Finance*, *37*, 337–342. https://doi.org/10.1016/S2212-5671(16)30134-4

Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces engage us: Photos with faces attract more likes and comments on Instagram. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 965–974).

Bakri, Z. F. (2023). Analyzing the influence of digital marketing strategies on business performance in the beauty industry: A comprehensive analysis of social media engagement and influencer collaborations. *Journal on Economics, Management and Business Technology*, *2*(1), 37–48. https://doi.org/10.35335/jembut.v2i1.187

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). *Everyone's an influencer: Quantifying influence on twitter* [Paper presentation]. Proceedings of the fourth ACM international conference on web search and data mining (pp. 65–74), New York, NY, USA. (Hong Kong, China) *(WSDM '11)*. Association for Computing Machinery. https://doi.org/10.1145/1935826.1935845

Balqis, V. A., Utami, S. G., Jessica, J., & Ikhsan, R. B. (2024). User-generated content by the beauty influencer and its impact on E-WOM and purchase intentions. In *2024 IEEE International conference on interdisciplinary approaches in technology and management for social innovation (IATMSI)* (Vol. 2, pp. 1–6). IEEE.

Barbosa dos Santos, M. L. (2022). The "so-called" UGC: An updated definition of user-generated content in the age of social media. *Online Information Review*, *46*(1), 95–113. https://doi.org/10.1108/OIR-06-2020-0258

Bazaco, Á., Redondo, M., & Sánchez-García, P. (2019). Clickbait as a strategy of viral journalism: Conceptualisation and methods. *Revista Latina de Comunicación Social*, *74*, 94–115. https://doi.org/10.4185/RLCS-2018-1323en.

Bazi, S., Filieri, R., & Gorton, M. (2023). Social media content aesthetic quality and customer engagement: The mediating role of entertainment and impacts on brand love and loyalty. *Journal of Business Research*, *160*, 113778. https://doi.org/10.1016/j.jbusres.2023.113778

Bennett, C. L., Gleason, C., Scheuerman, M. K., Bigham, J. P., Guo, A., & To, A. (2021). *"It's Complicated": Negotiating accessibility and (mis)representation in image descriptions of Race, gender, and disability* [Paper presentation]. Chi. Acm (pp. 375:1–375:19).

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1998). *Classification and regression trees*. Repr. *The Wadsworth statistics/probability series*. Wadsworth International Group. *c* 8, 8.

Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies*, *8*(2), 29–40.

BrydgesSjöholm, T. (2019). Becoming a personal style blogger: Changing configurations and spatialities of aesthetic labour in the fashion industry. *International Journal of Cultural Studies*, *22*(1), 119–139. https://doi.org/10.1177/1367877917752404

Caruana, R., Lawrence, S., & Giles, C. (2000). Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. In *Proceedings of the 13th Annual Conference on Neural Information Processing Systems (NeurIPS 2000)* (pp. 381–387).

Alves De Castro, C., & Carthy, A. (2021). *Social media influencers (SMIs) in context: A literature review*.

Chandrasekaran, S., Annamalai, B., & Kumar De, S. (2019). Evaluating marketer generated content popularity on brand fan pages—A multilevel modelling approach. *Telematics and Informatics*, *44*, 101266. https://doi.org/10.1016/j.tele.2019.101266

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), 16–28. 40th-year commemorative issue. https://doi.org/10.1016/j.compeleceng.2013.11.024

Chen, J. L., & Dermawan, A. (2020). The influence of YouTube beauty vloggers on Indonesian consumers' purchase intention of local cosmetic products. *International Journal of Business and Management*, *15*(5), 100–116. https://doi.org/10.5539/ijbm.v15n5p100

Cheong, H. J., & Morrison, M. A. (2008). Consumers' reliance on product information and recommendations found in UGC. *Journal of Interactive Advertising*, *8*(2), 38–49. https://doi.org/10.1080/15252019.2008.10722141

Chou, Y.-L., Lee, H.-J., Tsai, J., Liang, E.-C., & Chang, Y.-J. (2023). *I like their autonomy and closeness to me: Uncovering the perceived appeal of social-media influencers* [Paper presentation]. In Proceedings of the 2023 CHI conference on human factors in computing systems (p. 67419), (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA . https://doi.org/10.1145/3544548.3580898

Creator Academy. (2021). *How to create an attractive and high-quality cover and title*. https://www.xiaohongshu.com/explore/66c40c76000000001d0165c7?xsec_source=pc_creator&xsec_token=HB6qL26u905fnzoY77pKh14OUhgpAlFOi6T4chCF8p4D4%3D

Creator Academy. (2022). *Xiaohongshu finally tests cover PK feature i've been waiting for 3 years*. https://www.mad-men.com/articldetails/20015

Cui, G., Chung, S. Y.-H., Peng, L., & Wang, Q. (2024). Clicks for money: Predicting video views through a sentiment analysis of titles and thumbnails. *Journal of Business Research*, *183*, 114849. https://doi.org/10.1016/j.jbusres.2024.114849

De, A., & Lu, Z. (2024). *#PoetsOfInstagram: Navigating the practices and challenges of novice poets on Instagram* [Paper presentation]. Proceedings of the CHI conference on human factors in computing systems (p. 16216), New York, NY, USA. (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery. https://doi.org/10.1145/3613904.3642173

Reviglio della Venaria, U. (2020). *Personalization in social media: Challenges and opportunities for democratic societies* [Doctoral dissertation, Alma Mater Studiorum – Università di Bologna]. https://doi.org/10.6092/unibo/amsdottorato/9529

Dennhardt, S. (2014). *User-generated content and its impact on branding: How users and communities create and manage brands in social media* (pp. 1–129). https://doi.org/10.1007/978-3-658-02350-8

Ding, Z., Smith-Renner, A., Zhang, W., Tetreault, J. R., & Jaimes, A. (2023). Harnessing the power of LLMs: Evaluating human-AI text co-creation through the lens of news headline generation. In *EMNLP (Findings)* (pp. 3321–3339). Association for Computational Linguistics.

Elliot, A. J., & Maier, M. A. (2007). Color and psychological functioning. *Current Directions in Psychological Science*, *16*(5), 250–254. https://doi.org/10.1111/j.1467-8721.2007.00514.x

Elliot, A. J., & Maier, M. A. (2015). Color psychology: Effects of perceiving color on psychological functioning in humans. *Annual Review of Psychology*, *65*, 95–120. https://doi.org/10.1146/annurev-psych-010213-115035

Fan, Z., Chen, R., Hu, T., & Liu, Z. (2024). FairMT-Bench: Benchmarking fairness for multi-turn dialogue in conversational LLMs. *arXiv preprint arXiv:2410.19317*. https://arxiv.org/abs/2410.19317

Fink, B., Grammer, K., & Thornhill, R. (2001). Human (*Homo sapiens*) facial attractiveness in relation to skin texture and color. *Journal of Comparative Psychology (Washington, D.C.: 1983)*, *115*(1), 92–99. https://doi.org/10.1037/0735-7036.115.1.92

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414–2423).

Gero, K. I., Liu, V., & Chilton, L. (2022). *Sparks: Inspiration for science writing using language models* [Paper presentation]. In Proceedings of the 2022 ACM designing interactive systems conference (pp. 1002–1019), New York, NY, USA. (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery. https://doi.org/10.1145/3532106.3533533

Gligorea, I., Cioca, M., Oancea, R., Gorski, A.-T., Gorski, H., & Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. *Education Sciences*, *13*(12), 1216. https://doi.org/10.3390/educsci13121216

Gräve, J.-F., & Greff, A. (2018). Good KPI, good influencer? Evaluating success metrics for social media influencers. In *Proceedings of the 9th International Conference on Social Media and Society* (pp. 291–295).

Guo, Q., Zhou, S., Wu, Y., Peng, Z., & Ma, X. (2022a). *Understanding and modeling viewers' first impressions with images in online medical crowdfunding campaigns* [Paper presentation]. In Proceedings of the 2022 CHI conference on human factors in computing systems (p. 36120), New York, NY, USA. Article (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery. https://doi.org/10.1145/3491102.3501830

Guo, Q., Zhou, S., Wu, Y., Peng, Z., & Ma, X. (2022b). Understanding and modeling viewers' first impressions with images in online medical crowdfunding campaigns. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1–20).

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*, 1157–1182. https://doi.org/10.5555/944919.944968

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. arXiv:https://doi.org/10.1177/0008125620958166

Haralock, R. M., & Shapiro, L. G. (1991). *Computer and robot vision*. Addison-Wesley Longman Publishing Co., Inc.

Harrigan, P., Daly, T. M., Coussement, K., Lee, J. A., Soutar, G. N., & Evers, U. (2021). Identifying influencers on social media. *International Journal of Information Management*, *56*, 102246. https://doi.org/10.1016/j.ijinfomgt.2020.102246

Hübner Barcelos, R., Dantas, D. C., & Sénécal, S. (2018). Watch your tone: How a brand's tone of voice on social media influences consumer responses. *Journal of Interactive Marketing*, *41*(1), 60–80. https://doi.org/10.1016/j.intmar.2017.10.001

Hudders, L., & De Jans, S. (2022). Gender effects in influencer marketing: An experimental study on the efficacy of endorsements by same- vs. other-gender social media influencers on Instagram. *International Journal of Advertising*, *41*(1), 128–149. https://doi.org/10.1080/02650487.2021.1997455

Humphrey, S., Brown, S. M., Cross, S. J., & Mehta, R. (2021). Defining skin quality: Clinical relevance, terminology, and assessment. *Dermatologic Surgery*, *47*(7), 974–981. https://doi.org/10.1097/DSS.0000000000003079

Huschens, M., Briesch, M., Sobania, D., & Rothlauf, F. (2023). Do you trust ChatGPT? - Perceived credibility of human and AI-generated content. *arXiv preprint arXiv:2309.02524*. https://arxiv.org/abs/2309.02524

Jalali, N. Y., & Papatla, P. (2016). The palette that stands out: Color compositions of online curated visual UGC that attracts higher consumer interaction. *Quantitative Marketing and Economics*, *14*(4), 353–384. https://doi.org/10.1007/s11129-016-9178-1

Janssen, L., Schouten, A. P., & Croes, E. A. J. (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. https://doi.org/10.1080/02650487.2021.1994205

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Jin, S. V., & Ryu, E. (2020). "I'll buy what she's #wearing": The roles of envy toward and parasocial interaction with influencers in Instagram celebrity-based brand endorsement and social commerce. *Journal of Retailing and Consumer Services*, *55*, 102121. https://doi.org/10.1016/j.jretconser.2020.102121

Joo, J., Li, W., Steen, F. F., & Zhu, S.-C. (2014). Visual persuasion: Inferring communicative intents of images. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 216–223). IEEE. https://doi.org/10.1109/CVPR.2014.35

Kang, Y., Sun, Z., Wang, S., Huang, Z., Wu, Z., & Ma, X. (2021). MetaMap: Supporting visual metaphor ideation through multi-dimensional example-based exploration. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. https://doi.org/10.1145/3411764.3445325

Kim, D. Y., & Kim, H.-Y. (2021). Trust me, trust me not: A nuanced view of influencer marketing on social media. *Journal of Business Research*, *134*, 223–232. https://doi.org/10.1016/j.jbusres.2021.05.024

Kim, J. G., Kong, H. K., Karahalios, K., Fu, W.-T., & Hong, H. (2016). The power of collective endorsements: Credibility factors in medical crowdfunding campaigns. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 4538–4549).

Koh, B., & Cui, F. (2022). An exploration of the relation between the visual attributes of thumbnails and the view-through of videos: The case of branded video content. *Decision Support Systems*, *160*, 113820. https://doi.org/10.1016/j.dss.2022.113820

Kolmogorova, A., Kalinin, A., & Malikova, A. (2021). Semiotic function of empathy in text emotion assessment. *Biosemiotics*, *14*(2), 329–344. https://doi.org/10.1007/s12304-021-09434-y

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kumar, R. R., & Alok, K. (2020). Adoption of electric vehicle: A literature review and prospects for sustainability. *Journal of Cleaner Production*, *253*, 119911. https://doi.org/10.1016/j.jclepro.2019.119911

Kwak, D., Kim, H., & Lim, Y-k (2022). *Exploring users' experiences of "suggested posts" in social media through the lens of social networking and interactions* [Paper presentation]. In *Companion publication of the 2022 conference on computer supported cooperative work and social computing* (pp. 37–40).

Labrecque, L. I., & Milne, G. R. (2012). Exciting red and competent blue: The importance of color in marketing. *Journal of the Academy of Marketing Science*, *40*(5), 711–727. https://doi.org/10.1007/s11747-010-0245-y

Lee, M. T., & Theokary, C. (2021). The superstar social media influencer: Exploiting linguistic style and emotional contagion over content? *Journal of Business Research*, *132*, 860–871. https://doi.org/10.1016/j.jbusres.2020.11.014

Li, Q., Huang, Z. J., & Christianson, K. (2016). Visual attention toward tourism photographs with text: An eye-tracking study. *Tourism Management*, *54*(2016), 243–258. https://doi.org/10.1016/j.tourman.2015.11.017

Li, R., Wu, C.-H., Liu, S., Wang, J., Wang, G., Liu, G., & Zeng, B. (2021). SDP-GAN: Saliency detail preservation generative adversarial networks for high perceptual quality style transfer. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, *30*(2021), 374–385. https://doi.org/10.1109/TIP.2020.3036754

Li, S., Wu, S., Liu, T., Zhang, H., Guo, Q., & Peng, Z. (2024). *Understanding the features of text-image posts and their received social support in online grief support communities* [Paper presentation]. Icwsm (pp. 917–929). AAAI Press.

Lindsey-Mullikin, J., & Borin, N. (2017). Why strategy is key for successful social media sales. *Business Horizons*, *60*(4), 473–482. https://doi.org/10.1016/j.bushor.2017.03.005

Liu, C., Huang, Z., Liu, D., Zhou, S., Peng, Z., & Ma, X. (2022). PlanHelper: Supporting activity plan construction with answer posts in community-based QA platforms. *Proceedings of the ACM on Human–Computer Interaction*, *6*(CSCW2), 1–26. https://doi.org/10.1145/3555555

Liu, C., Zhou, S., Liu, D., Li, J., Huang, Z., & Ma, X. (2023). *CoArgue : Fostering Lurkers' contribution to collective arguments in community-based QA platforms* [Paper presentation]. Proceedings of the 2023 CHI conference on human factors in computing systems (p. 27117), New York, NY, USA. (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery. https://doi.org/10.1145/3544548.3580932

Liu, J., Lughofer, E., & Zeng, X. (2015). Aesthetic perception of visual textures: A holistic exploration using texture analysis, psychological experiment, and perception modeling. *Frontiers in Computational Neuroscience*, *9*, 134. https://doi.org/10.3389/fncom.2015.00134

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. https://doi.org/10.1080/15252019.2018.1533501

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. *Advances in Neural Information Processing Systems*, *26*, 431–439.

Lyu, Y., Zhang, H., Niu, S., & Cai, J. (2024). *A preliminary exploration of Youtubers' use of Generative-AI in content creation* [Paper presentation]. *CHI extended abstracts* (pp. 20:1–20:7). ACM.

Ma, R., Gui, X., & Kou, Y. (2023). Multi-platform content creation: The configuration of creator ecology through platform prioritization, content synchronization, and audience management. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1–19).

Ma, Z., Dou, Z., Zhu, Y., Zhong, H., & Wen, J.-R. (2021). *One Chatbot per person: Creating personalized Chatbots based on implicit user profiles* [Paper presentation]. Sigir (pp. 555–564). ACM.

Machajdik, J., & Hanbury, A. (2010). Affective image classification using features inspired by psychology and art theory. In *Proceedings of the 18th ACM international conference on multimedia* (pp. 83–92).

Mahmood, A., & Huang, C.-M. (2022). *Effects of rhetorical strategies and skin tones on agent persuasiveness in assisted decision-making* [Paper presentation]. Proceedings of the 22nd ACM international conference on intelligent virtual agents (p. 78), New York, NY, USA. Article (Faro, Portugal) *(IVA '22)*. Association for Computing Machinery. https://doi.org/10.1145/3514197.3549628

Malik, A., & Srinivasam, S. (2020). A comparative analysis of various social networking sites in marketing. In *2020 2nd International conference on advances in computing, communication control and networking (ICACCCN)* (pp. 26–31). IEEE.

Mangold, W. G., & Faulds, D. J. (2009). Social media: The new hybrid element of the promotion mix. *Business Horizons*, *52*(4), 357–365. https://doi.org/10.1016/j.bushor.2009.03.002

N. J., Lee, M. D., Nayottama A.p, Anderies., & A., Chowanda. (2023). How digitally extractable attributes of YouTube video thumbnails and titles affect video views. In *2023 IEEE 7th international conference on information technology, information systems and electrical engineering (ICITISEE)* (pp. 82–87). https://doi.org/10.1109/ICITISEE58992.2023.10405243

Nuno Emanuel Branquinho Moutinho Marques de Paiva. (2018). *Using emoji in an e-commerce context: Effects in brand perception, quality of service and intention to recommend* [Master's thesis].

Oh, C., Choi, J., Lee, S., Park, S., Kim, D., Song, J., Kim, D., Lee, J., & Suh, B. (2020). *Understanding user perception of automated news generation system* [Paper presentation]. Chi (pp. 1–13). ACM.

Olivier, P., Halper, N., Pickering, J., & Luna, P. (1999). *Visual composition as optimisation* [Paper presentation]. AISB symposium on AI and creativity in entertainment and visual art (Vol. 1, pp. 22–30). Newcastle University.

Osgood, C. E. (1957). *The measurement of meaning*. Universit of Illinois.

Park, J. Y., Jang, J., Jaimes, A., Chung, C.-W., & Myaeng, S.-H. (2014). *Exploring the user-generated content (UGC) uploading behavior on YouTube* [Paper presentation]. Proceedings of the 23rd International conference on World Wide Web (pp. 529–534).

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, *32*, 8024–8035.

Pechová, M., Lohr, V., & Havlíček, Z. (2015). Social media for organic products promotion. *Agris on-Line Papers in Economics and Informatics*, *7*(1), 41–50. https://doi.org/10.7160/aol.2015.070105

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., Vincent, M., Joly, A., Stelzer, M., … Müller, A. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Peng, L., Cui, G., Chung, Y., & Zheng, W. (2020a). The faces of success: Beauty and ugliness premiums in e-commerce platforms. *Journal of Marketing*, *84*(4), 67–85. https://doi.org/10.1177/0022242920914861

Peng, Z., Chen, Q., Shen, Z., Ma, X., & Oulasvirta, A. (2024). DesignQuizzer: A community-powered conversational agent for learning visual design. *Proceedings of the ACM on Human–Computer Interaction*, *8*(CSCW1), 1–40. https://doi.org/10.1145/3637321

Peng, Z., Guo, Q., Tsang, K. W., & Ma, X. (2020b). *Exploring the effects of technological writing assistance for support providers in online mental health community* [Paper presentation]. Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1–15), New York, NY, USA. (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery. https://doi.org/10.1145/3313831.3376695

Peng, Z., Ma, X., Yang, D., Tsang, K. W., & Guo, Q. (2021). *Effects of support-seekers' community knowledge on their expressed satisfaction with the received comments in mental health communities* [Paper presentation]. Proceedings of the 2021 CHI conference on human factors in computing systems (p. 53612), New York, NY, USA. Article (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery. https://doi.org/10.1145/3411764.3445446

Pias, S. B. H., Huang, R., Williamson, D. S., Kim, M., & Kapadia, A. (2024). *The impact of perceived tone, age, and gender on voice assistant persuasiveness in the context of product recommendations* [Paper presentation]. Proceedings of the 6th ACM conference on conversational user interfaces (p. 2015), New York, NY, USA. Article (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery. https://doi.org/10.1145/3640794.3665545

Pieters, R., & Wedel, M. (2004). Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *Journal of Marketing*, *68*(2), 36–50. https://doi.org/10.1509/jmkg.68.2.36.27794

Porcheron, A., Latreille, J., Jdid, R., Tschachler, E., & Morizot, F. (2014). Influence of skin ageing features on Chinese women's perception of facial age and attractiveness. *International Journal of Cosmetic Science*, *36*(4), 312–320. https://doi.org/10.1111/ics.12128

Qian, H., Dou, Z., Zhu, Y., Ma, Y., & Wen, J.-R. (2021). Learning implicit user profile for personalized retrieval-based Chatbot. *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21)* (pp. 1467–1477). https://doi.org/10.1145/3459637.3482269

Qin, H. X., Jin, S., Gao, Z., Fan, M., & Hui, P. (2024a). *CharacterMeet: Supporting creative writers' entire story character construction processes through conversation with LLM-powered chatbot avatars* [Paper presentation]. Chi (pp. 1051:1–1051:19). ACM.

Qin, M., Qiu, S., Zhao, Y., Zhu, W., & Li, S. (2024b). Graphic or short video? The influence mechanism of UGC types on consumers' purchase intention—Take Xiaohongshu as an example. *Electronic Commerce Research and Applications*, *65*, 101402. https://doi.org/10.1016/j.elerap.2024.101402

Ramadhan, H. A. R., Yeni, T., Misbak, M., & Muthiarsih, T. (2025). TikTok as a marketing medium: The influence of user generated content (UGC) and viral marketing on online purchase decisions on Glad2Glow skincare products. *International Journal of Business, Economics, and Social Development*, *6*(2), 282–291. https://doi.org/10.46336/ijbesd.v6i2.945

Ren, J., Raghupathi, V., & Raghupathi, W. (2020). Understanding the dimensions of medical crowdfunding: A visual analytics approach. *Journal of Medical Internet Research*, *22*(7), e18813. https://doi.org/10.2196/18813

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception*, *35*(6), 749–759. https://doi.org/10.1068/p5490

Shahbaznezhad, H., Dolan, R., & Rashidirad, M. (2021). The role of social media content format and platform in users' engagement behavior. *Journal of Interactive Marketing*, *53*(1), 47–65. https://doi.org/10.1016/j.intmar.2020.05.001

Sharma, N., Singh, A., & Vashisht, A. (2024). *Influencer marketing on new platforms and the metaverse for male grooming brands* (pp. 197–213). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65727-6_12

Shin, D., He, S., Lee, G. M., Whinston, A., Cetintas, S., & Lee, K. C. (2020). Enhancing social media analysis with visual data analytics: A deep learning approach. *Management Information Systems Quarterly*, *44*(4), 1459–1492.

Singh Tanwar, A., Chaudhry, H., & Srivastava, M. K. (2022). Trends in influencer marketing: A review and bibliometric analysis. *Journal of Interactive Advertising*, *22*(1), 1–27. https://doi.org/10.1080/15252019.2021.2007822

Singh, K. (2021). Influencer marketing from a consumer perspective: How attitude, trust, and word of mouth affect buying behavior. *European Integration Studies*, *1*(15), 231–241. https://doi.org/10.5755/j01.eis.1.15.28803

Steele, F., Jr, Evans, D., & Green, R. (2009). Is your profile picture worth 1000 words? Photo characteristics associated with personality impression agreement. In *Proceedings of the International AAAI conference on web and social media* (Vol. 3, pp. 327–330).

Tafesse, W., & Wood, B. P. (2021). Followers' engagement with Instagram influencers: The role of influencers' content and engagement strategy. *Journal of Retailing and Consumer Services*, *58*, 102303. https://doi.org/10.1016/j.jretconser.2020.102303

Tan, K. W., Tiddeman, B., & Stephen, I. D. (2018). Skin texture and colour predict perceived health in Asian faces. *Evolution and Human Behavior*, *39*(3), 320–335. https://doi.org/10.1016/j.evolhumbehav.2018.02.003

Thomas, R. J., Masthoff, J., & Oren, N. (2019). Can i influence you? Development of a scale to measure perceived persuasiveness and two studies showing the use of the scale. *Frontiers in Artificial Intelligence*, *2*(2019), 24. https://doi.org/10.3389/frai.2019.00024

Truong, A., Chi, P., Salesin, D., Essa, I., & Agrawala, M. (2021). Automatic generation of two-level hierarchical tutorials from instructional makeup videos. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–16).

Van De Weijer, J., Schmid, C., & Verbeek, J. (2007). *Learning color names from real-world images* [Paper presentation]. 2007 IEEE conference on computer vision and pattern recognition (pp. 1–8). IEEE.

Wadhwa, M., & Zhang, K. (2015). This number just feels right: The impact of roundedness of price numbers on product evaluations. *Journal of Consumer Research*, *41*(5), 1172–1185. https://doi.org/10.1086/678484

Weber, P., Ludwig, T., Brodesser, S., & Grönewald, L. (2021). *"It's a Kind of Art!": Understanding food influencers as influential content creators* [Paper presentation]. Proceedings of the 2021 CHI conference on human factors in computing systems (pp. 17514), New York, NY, USA. Article (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery. https://doi.org/10.1145/3411764.3445607

Wei, X., & Stillwell, D. (2017). How smart does your profile image look?: Estimating intelligence from social network profile images. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (WSDM '17)* (pp. 33–40). Association for Computing Machinery. https://doi.org/10.1145/3018661.3018663

Woolson, R. F. (2008). Wilcoxon Signed-Rank Test. 3. https://doi.org/10.1002/9780471462422.eoct979

Wu, Z., Kim, T., Li, Q., & Ma, X. (2019). *Understanding and modeling user-perceived brand personality from mobile application UIs* [Paper presentation]. Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1–12), New York, NY, USA. (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery. https://doi.org/10.1145/3290605.3300443

Xie, S., Wei, H., & Liu, F. (2023). Is beauty always good? Effects of visual presentation of Influencer's aesthetic labor on brand purchase intention. *Journal of Retailing and Consumer Services*, *75*, 103528. https://doi.org/10.1016/j.jretconser.2023.103528

Yacouby, R., & Axman, D. (2020). Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In *Proceedings of the first workshop on evaluation and comparison of NLP systems* (pp. 79–91).

Yang, H., Yuan, J., Yang, S., Xu, L., Yuan, S., & Zeng, Y. (2024). A new creative generation pipeline for click-through rate with stable diffusion model. In *Companion proceedings of the ACM Web conference 2024 (WWW '24)* (pp. 180–189). Association for Computing Machinery. https://doi.org/10.1145/3589335.3648315

Yu, C.-E., Xie, S. Y., & Wen, J. (2020). Coloring the destination: The role of color psychology on Instagram. *Tourism Management*, *80*(2020), 104110. https://doi.org/10.1016/j.tourman.2020.104110

Zafar, A. U., Qiu, J., Li, Y., Wang, J., & Shahzad, M. (2021). The impact of social media celebrities' posts and contextual interactions on impulse buying in social commerce. *Computers in Human Behavior*, *115*(2021), 106178. https://doi.org/10.1016/j.chb.2019.106178

Zhang, S., Lee, D., Singh, P., & Srinivasan, K. (2021a). What makes a good image? Airbnb demand analytics leveraging interpretable image features. *Management Science*, *68*(8), 5644–5666. https://doi.org/10.1287/mnsc.2021.4175

Zhang, X., Lyu, H., & Luo, J. (2021b). What contributes to a crowdfunding campaign's success? Evidence and analyses from GoFundMe data. *Journal of Social Computing*, *2*(2), 183–192. https://doi.org/10.23919/JSC.2021.0010

Zhang, Y., & Pennacchiotti, M. (2013). *Predicting purchase behaviors from social media* [Paper presentation]. In Proceedings of the 22nd international conference on World Wide Web (pp. 1521–1532), New York, NY, USA. (Rio de Janeiro, Brazil) *(WWW '13)*. Association for Computing Machinery. https://doi.org/10.1145/2488388.2488521

Zhang, Z., Gao, J., Singh Dhaliwal, R., & Li, T. J.-J. (2023). *VISAR: A human-AI argumentative writing assistant with visual programming and rapid draft prototyping* [Paper presentation]. Proceedings of the 36th Annual ACM Symposium on user interface software and technology (p. 530), New York, NY, USA. Article (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery. https://doi.org/10.1145/3586183.3606800

Zhao, J., Sha, B., & Deng, S. (2024). More repeated exposures, lower preference ratings: Exploring the factors impacting on visual aesthetic fatigue. *Environment, Development and Sustainability*. https://doi.org/10.1007/s10668-024-05683-2

Zhao, Z., Chen, X., Yang, J., & Jun, Z. (2022). How do the visual and textual factors influence user's attitudes? An elaboration likelihood model perspective. In *Pacific Asia Conference on Information Systems (PACIS) 2022 Proceedings* (p. 130).

Zhou, K. Z., Shen, B., Zimmer, F., Xia, C., & Tong, X. (2023). *More than a wife and a mom: A study of mom vlogging practices in China* [Paper presentation]. CSCW companion (pp. 56–63). ACM.

Zhu, T. (2020). Analysis on the applicability of the random forest. *Journal of Physics: Conference Series*, 1607, 012123.

Zia, M., Zahra, M. F., & Hayat, N. (2021). Instagram beauty influencers and purchase decisions: Exploring the mediating role of source credibility. *VFAST Transactions on Education and Social Sciences*, *9*(3), 01–10.

Zuhroh, N. A., & Rakhmawati, N. A. (2020). Clickbait detection: A literature review of the methods used. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, *6*(1), 1–10. https://doi.org/10.26594/register.v6i1.1561

About the authors

Xueer Lin is a Master's student at the School of Artificial Intelligence, Sun Yat-sen University. Her research focuses on Human–Computer Interaction (HCI), with an emphasis on social media applications and user behavior.

Xia Chen is an incoming graduate student at the School of Artificial Intelligence, Sun Yat-sen University. Her research interests focus on Human–Computer Interaction (HCI), aiming to improve practical applications that enhance how humans interact with technology.

Zhenhui Peng is currently working as an associate professor in the School of Artificial Intelligence at Sun Yat-sen University. With a PhD from HKUST, his research interests are in Human–Computer Interaction (HCI) and Social Computing, particularly in building intelligent systems to support users in higher-order thinking tasks.

Appendix A: Order effects of experienced tool and task in the within-subjects study

Table 6. The *p* values of the mixed-ANOVA tests to examine the impact of tool/task orders.

	Within-subjects effects				Between-subjects effects		
	Tool	Tool * order	Tool * task	Tool * order * task	Order	Task	Order * task
Sum	<0.001	0.906	0.483	0.108	0.449	0.754	0.351
-Persuasiveness	<0.001	0.849	0.485	0.656	0.577	0.966	0.466
-Relevance	<0.001	0.670	0.523	0.290	0.868	0.509	0.741
-Trustworthiness	0.006	0.869	0.299	0.052	0.261	0.805	0.304
Mental demand	0.036	0.070	0.607	0.968	0.442	0.971	0.486
Physical demand	<0.001	0.173	0.173	0.408	0.032	0.136	0.532
Temporal demand	<0.001	0.167	0.167	0.726	0.189	**0.012**	0.739
Own Performance	<0.001	<0.001	0.197	0.386	0.093	0.159	**0.015**
Effort	<0.001	**0.035**	0.165	0.754	0.164	0.119	0.164
Frustration	<0.001	0.216	0.301	0.677	0.072	0.357	0.711
System Usability Scale	<0.001	0.275	0.821	0.622	0.184	0.282	0.925
-Effectiveness & learnability	<0.001	0.773	0.828	0.566	0.387	0.223	0.999
-Efficiency in use	<0.001	0.241	0.578	0.577	**0.038**	0.099	0.586
-Satisfaction	<0.001	0.242	0.735	1.000	0.098	0.865	0.400
Vigilance	<0.001	0.196	0.471	1.000	**0.042**	0.295	0.620
Potential Negative Impact	0.001	**0.039**	0.239	0.812	0.030	0.451	0.451
Trust in Ethical Standards	0.032	0.804	**0.032**	0.804	0.501	0.822	1.000
Commitment to Users	<0.001	0.075	0.389	0.216	0.180	0.466	0.754
Reliability of Outputs	<0.001	0.068	0.406	0.532	0.462	0.329	0.806

Within-subjects: **tool** (*BeautyClicker* vs. baseline); between-subjects: the **order** of the used tool, the order of learning **task**; N = 32.

Appendix B: Detailed information of 32 participants of within-subjects study

Table 7. Detailed demographic and response summary of the 32 participants in the within-subjects study.

User	Gender	Age	Occupation	Major	Post experience	Beauty post experience	Beauty browsing frequency	Cover design experience	Title and content writing experience	LLM usage experience	AI tool interest
1	Female	24	Undergraduate	Artificial Intelligence	4	2	Almost every day	2	4	6	7
2	Female	19	Undergraduate	Artificial Intelligence	4	1	Almost every day	5	4	6	5
3	Female	19	Undergraduate	Artificial Intelligence	4	2	Almost every day	4	4	7	7
4	Female	22	Undergraduate	Geophysics	3	2	Almost every day	2	2	7	6
5	Female	22	Undergraduate	Geology	3	1	Almost every day	1	3	5	5
6	Female	23	Undergraduate	Computer Science	6	4	Almost every day	4	4	7	5
7	Female	22	Undergraduate	Artificial Intelligence	4	3	2-6 times per week	3	4	7	7
8	Female	21	Undergraduate	Artificial Intelligence	3	1	2-6 times per week	1	2	1	4
9	Female	24	Undergraduate	Computer Science	4	4	2-6 times per week	2	3	7	7
10	Female	22	Undergraduate	Computer Science	5	1	2-6 times per week	3	5	7	7
11	Female	22	Undergraduate	Computer Science	4	2	2-6 times per week	3	5	6	7
12	Female	22	Graduate	Computer Science	6	4	Almost every day	4	4	7	7
13	Female	22	Undergraduate	Computer Technology	5	5	Almost every day	2	4	4	6
14	Female	20	Undergraduate	Ocean Engineering	2	1	Almost every day	1	2	5	7
15	Female	22	Undergraduate	Chemical Engineering	6	6	Almost every day	6	6	6	6
16	Female	22	Undergraduate	Artificial Intelligence	5	4	Almost every day	5	6	7	4
17	Female	21	Graduate	Material Chemistry	6	5	Once a week	5	5	7	6
18	Female	23	Graduate	Computer Science	5	2	Almost every day	5	5	7	6
19	Female	21	Graduate	Artificial Intelligence	4	1	2-6 times per week	3	4	7	6
20	Female	21	Undergraduate	Chemical Engineering	2	2	Almost every day	2	2	2	7
21	Female	20	Graduate	Korean Language	4	1	2-6 times per week	2	4	4	4
22	Female	22	Employee	nan	7	5	Almost every day	5	6	6	7
23	Female	22	Employee	nan	3	3	2-6 times per week	3	3	4	5
24	Female	21	Graduate	Philosophy	2	1	2-6 times per week	3	2	4	4
25	Female	20	Graduate	Artificial Intelligence	1	1	2-6 times per week	1	1	5	3
26	Female	21	Undergraduate	Water Resources Engineering	5	1	Almost every day	6	6	6	7
27	Female	22	Graduate	Microelectronics	1	1	2-6 times per week	1	2	4	5
28	Female	24	Graduate	Energy Power Engineering	3	1	Once a week	2	2	3	6
29	Female	22	Graduate	Computer Technology	4	2	Almost every day	2	3	6	5
30	Female	22	Graduate	Artificial Intelligence	5	4	Almost every day	2	4	7	7
31	Female	22	Graduate	Artificial Intelligence	4	1	Almost every day	1	1	6	7
32	Female	22	Graduate	E-Commerce and Internet Computing	5	2	2-6 times per week	3	3	4	5

Each participant responded to a series of experience-related questions on a 1–7 likert scale (1 = no experience, 7 = very extensive experience), covering: **Post experience** *("how much experience do you have posting image-text content on platforms such as xiaohongshu?")*, **beauty post experience** *("how much experience do you have posting beauty-related notes?")*, **cover design experience** *("how much experience do you have designing cover images?")*, **title and content writing experience** *("how much experience do you have writing post titles and captions?")*, **LLM usage experience** *("how much experience do you have using large language models like ChatGPT or wenxin yiyan?")*, and **AI tool interest** *("how interested are you in using online resources and AI-assisted tools for cover design and writing?")*. the **beauty browsing frequency** column reflects the self-reported frequency with which participants browse beauty-related content in RED.