Competing for Attention: Text Overlays and Showing Faces in Nonprofit Instagram Posts

Sofia PANSONI, Università di Macerata, s.pansoni@unimc.it
Giacomo GISTRI, Università di Macerata, giacomo.gistri@unimc.i

Abstract

Nonprofit organizations increasingly rely on visually rich social media posts to capture attention and mobilize support. Among emerging design practices, text overlays (e.g., words or phrases deliberately inserted within an image) are gaining prominence, yet their interaction with other salient visual cues remains underexplored. This study investigates how text overlays influence engagement and whether this effect depends on the simultaneous presence of human faces depicted in the image. To address this gap, we analyzed 1982 Instagram posts published between April 2023 and June 2025 by UNICEF International, Save the Children US, and Doctors Without Borders. Images were annotated through a computational content analysis approach powered by generative artificial intelligence, which simultaneously detected text overlays and face presence.

The findings reveal that text overlays significantly increase engagement, but this positive effect disappears when human faces are present, indicating that the impact of textual cues depends on the surrounding visual context. These results advance understanding of multimodal nonprofit communication and provide practical guidance for designing more effective social media appeals.

Keywords: *Charity campaigns; Social media engagement; Text overlays; Computational Content Analysis*

Introduction

In the last decade, social media platforms became vital tools for nonprofit organizations (NPOs) to improve their outreach, interact with their supporters, and raise financial support in a more and more saturated digital space. Entities like the United Nations Development Programme (UNDP) characterize this setting as a rapidly evolving, densely populated sphere where attention is ephemeral, necessitating constant adaptation (UNDP, 2024). The 2025 M+R Benchmarks Study reaffirms the importance of these platforms for fundraising and advocacy, noting that 99% of NPOs maintain a presence on Facebook and 98% on Instagram, the latter exhibiting a notable audience increase (+11% in 2024). In this context, visual storytelling has emerged as an essential strategy to capture attention and cultivate emotional involvement. One significant factor influencing this phenomenon is how imagery can engage viewers in stories that resonate emotionally. Literature on charity advertising largely focuses on the role of human faces, particularly those belonging to recognizable beneficiaries, in eliciting empathy, enhancing engagement, and encouraging donations (Small & Loewenstein, 2003; Small & Verrochi, 2009; Nikulina et al., 2024). However, the performance of facial imagery in nonprofit campaigns is strongly linked to context and is shaped by diverse visual and textual design features. Among these features, the use of text overlays (TO) has rapidly emerged as a distinctive social media trend. NPOs increasingly superimpose short messages, calls-to-action, or slogans directly onto images, leveraging TO as a way to break through the clutter of fast-scrolling feeds (Farace et al. 2025; Wooley et al. 2022). Unlike captions that necessitate additional cognitive effort from users, TO provides immediate semantic cues, thereby augmenting the clarity of appeals (Farace et al., 2025). By simplifying key information into visually impactful text features, nonprofits can create a sense of urgency, spotlight campaign phrases, or elevate the emotional weight of visuals. Despite this, there is a notable lack of understanding about the way TO interact with various visual elements of a social media post, particularly faces, in shaping engagement levels. Drawing on theories of visual salience, as well as recent research on multimodal communication design, we investigate whether TO enhances or diminishes the persuasive impact of faces on audience engagement. Our findings try to contribute to the literature on charitable communication by offering innovative perspectives on how visual and textual configurations affect audience reactions, as well as practical strategies for nonprofits to better their communication efforts in an increasingly fierce online marketplace.

Theoretical Background

On fast-moving social media feeds, where countless images compete for users' attention, nonprofit organizations need visual strategies that instantly communicate their appeals. Text overlays, defined as "*words or phrases (beyond a brand trademark or logo) that content providers insert deliberately within a delimited area of an image in a social media post*" (Farace et al., 2025), have emerged as an increasingly adopted practice. By embedding concise textual cues inside the visual field, TO allow audiences to grasp the intended message even with a fleeting glance, in contrast to captions that require more deliberate reading (Wooley & Sharif, 2022). Theoretically, TO can increase engagement through several mechanisms. They provide immediate semantic anchoring, lowering cognitive effort and enhancing message clarity (Farace et al., 2025; Pieters &Wedel, 2004). They also create visual salience that helps posts stand out from competing stimuli (Arnheim, 1954; Pieters &Wedel, 2004), and they integrate verbal and visual cues into a unified composition. We therefore propose the following hypothesis:

H1: *Social media posts containing a text overlay are expected to generate higher engagement than posts without a text overlay.*

While TO can act as a visual accelerator of meaning, its effectiveness may depend on how it interacts with other salient visual elements. Notably, human faces are among the most powerful visual cues in charitable communication because they elicit empathy and identification (Small & Verrochi, 2009; Nikulina et al., 2024). However, prior research also highlights important contextual factors. Overly negative facial expressions can induce discomfort or skepticism (Kwon et al., 2021), and racial or gendered cues can activate stereotypes that reduce donations (Yazdani, Chakravarty, & Inman, 2025). Moreover, when multiple focal elements compete for viewers 'attention, visual clutter can occur, diminishing overall appeal (Pieters, Wedel, & Zhang, 2007). Farace et al. (2025) emphasize that the specific design of the text overlay, such as its size, position, and contrast, plays a decisive role: overly salient configurations can create visual imbalance and reduce engagement when combined with already prominent image elements. Human faces, especially expressive ones, are intrinsically salient stimuli that naturally attract attention (Henderson et al. 1999). Building on visual salience and aesthetic processing theories (Locher et al. 1999; Affonso & Janiszewski, 2023), we argue that faces and TO can either reinforce or compete with each other. When no face is present, TO can take center stage and guide viewers' understanding. When a face is present, however, a TO may split attention, reduce perceptual fluency, and weaken the positive effect of facial cues on engagement.
Accordingly, we posit the following hypothesis:

H2. The effect of text overlay on engagement is moderated by the presence of human faces: when no face is present, text overlay increases engagement, whereas when a face is present, text overlay decreases engagement.

Methodology
To test our hypothesis, we collected 1982 Instagram posts between April 2023 and June 2025. Posts were scraped using the Apify platform from three top-performing international nonprofit organizations: UNICEF International, Save the Children US, and Doctors Without Borders. These accounts were selected based on three criteria: First, we focused on organizations that communicate primarily in English. This choice allowed us to rely on validated natural language processing (NLP) dictionaries and to perform multimodal annotation through GPT-based APIs, while also ensuring semantic consistency in the interpretation of both textual and visual content. Second, we selected accounts with a large followers base (≥ 1 million). This threshold identified leading actors in the nonprofit sector, characterized by high visibility and substantial global influence. Finally, we selected organizations with high levels of social media activity. In particular, we included accounts that had published more than 2,000 posts (images and videos) and had maintained a steady posting rhythm since January 1, 2023. This ensured access to a rich and up-to-date dataset, reflecting current practices rather than historical patterns of communication.
A distinctive methodological contribution of this study lies in the AI-driven computational content analysis (CCA) adopted to annotate visual and textual features of the collected images. As highlighted by Chang (2017), traditional content analysis is constrained by heavy manual effort, risk of human error, and the tendency to reduce multimedia complexity. Data science and machine learning now make it possible to examine large-scale multimodal datasets, encompassing text, images, audio, and video, without such simplifications (Huh & Malthouse, 2020; Berger et al., 2020). Recent applications demonstrate the potential to automatically detect objects, emotions, and themes, or to conduct topic modeling and clustering at scale, revealing patterns invisible to manual coding (Ordenes et al., 2019; Shi et al., 2023). Nevertheless, AI-driven CCA has limitations. Supervised approaches typically require large annotated datasets that are costly and time-consuming

to create, and fully automated systems may uncover statistical regularities without providing theoretical clarity (Barari et al. 2024). Moreover, most applications still focus on unimodal content, especially text, while contemporary advertising increasingly relies on multimodal communication (Barari et al. 2024).

Against this backdrop, our study introduces an innovative prompt-based multimodal procedure. Rather than combining separate models for facial recognition and optical character recognition (OCR), we rely on a single large language model (ChatGPT 4.1) capable of simultaneously detecting human faces and text overlay.

The annotation process involved a prompt engineering phase, during which the prompt was iteratively refined and tested in the ChatGPT Playground. The Playground is an interactive web-based environment provided by OpenAI that allows users to design, test, and refine prompts in real time. This tool enabled us to optimize the wording and structure of the prompt, verifying the accuracy of model outputs before automating the procedure through Python. Once validated, the procedure was fully automated through a custom Python script. The script sequentially encoded images, submitted them to the API together with the annotation prompt, and stored the structured outputs in JSONL format. To ensure reliability, a validation check was performed: a random subsample of 50 images was manually coded by an independent human coder. The comparison showed full consistency between automated and manual annotations, providing strong evidence for the accuracy of the automated procedure.

Building on these annotated data, we defined the variables for the empirical analysis. The dependent variable is engagement, measured as the logarithm of the total number of interactions (likes and comments) on each post. The log transformation was applied to reduce skewness and approximate normality. The focal independent variable is TO presence, coded 0 when no text overlay is present and 1 when it is present. The moderator is Face presence, coded 0 when no human face is visible and 1 when at least one face is shown. The moderation hypothesis was tested using Hayes' PROCESS macro for SPSS (Model 1; Hayes, 2022). In this specification, TO presence was entered as the independent variable (X), engagement as the dependent variable (Y), and face presence as the moderator (W). The model estimated the conditional effect of TO on engagement at different levels of the moderator (Face = 0 vs. Face = 1). Standard errors were computed using 5,000 bootstrap resamples, with 95% bias-corrected confidence intervals. The significance of the interaction term (TO × Face presence) was evaluated to determine whether the presence of faces moderates the effect of text overlay on engagement.

Results

To test the hypothesized moderation model, we used Hayes' PROCESS macro for SPSS (Model 1) with 5,000 bootstrapped samples (Hayes, 2018; Zhao, Lynch, & Chen, 2010). The dependent variable was engagement, measured as the log-transformed number of likes and comments. The overall model was significant ($R = 0.47$; $R^2 = 0.23$; $F(6,1975) = 95.65$, $p < 0.001$). Results show that TO presence positively affects engagement ($b = 0.24$, $p < 0.05$), supporting H1. Table 1 presents the complete PROCESS Model 1 results. As predicted in H2, face presence significantly moderates the effect of TO on engagement ($b = –0.39$, $p = 0.002$), indicating that the impact of TO depends on whether a human face appears in the post. The conditional effects are summarized in Table 2. Specifically, when no face was present, TO presence significantly increased engagement ($b = 0.24$, $p = 0.03$). Conversely, when a face was present, TO presence significantly decreased engagement ($b = –0.15$, $p = 0.01$). These findings confirm that text overlay boosts engagement only in posts without human faces and exerts a reverse (negative) effect when faces are included. Among the covariates, log follower count and color concentration displayed positive and significant relationships with engagement, whereas weekend posting did not reach significance.

Predictor	B	SE	t	*p*	95% CI
Constant	1.46	0.37	3.92	.000	[0.73, 2.19]
TO presence (x)	0.24	0.11	2.13	.033	[0.02, 0.46]
Face presence (w)	–0.16	0.10	–1.58	.113	[–0.37, 0.04]
TO × face presence	–0.39	0.12	–3.11	.002	[–0.63, –0.14]
Logarithm of followers	0.43	0.02	18.47	<.001	[0.38, 0.47]
Weekend posting	0.08	0.05	1.60	.110	[–0.02, 0.17]
Color concentration	0.007	0.001	6.50	<.001	[0.005, 0.009]

Table 1: Regression coefficients from PROCESS Model 1 - $R^2 = 0.23$; $F(6,1975) = 95.65$, $p < 0.001$.

.

Face presence	*Effect of TO on Engagement*	*SE*	*t*	*p*	*95% CI*
Absent	0.24	0.11	2.13	.033	[0.02, 0.46]
Present	–0.15	0.06	–2.44	.015	[–0.26, –0.03]

Table 2: Conditional effects of TO presence on engagement at different levels of face presence.

Conclusions and Future Development

From a theoretical perspective, our findings enrich the literature on the interplay between textual and visual elements in nonprofit communication. While faces are typically used to elicit empathy and foster social connection, they seem to alter the positive effect of text overlay. A plausible explanation is that faces and TO compete for limited visual and emotional attention, reducing perceptual fluency and weakening the persuasive effect of textual cues. In doing so, this study complements and extends prior work on visual salience and multimodal message design, offering new evidence that the effectiveness of textual elements is contingent on the surrounding visual configuration.

From a managerial standpoint, the results highlight the need for organizations to carefully align text overlay with visual composition: the inclusion of textual elements is beneficial when posts focus on symbolic or contextual imagery, where it can guide viewers' understanding and convey urgency. By contrast, it may be counterproductive when combined with the display of beneficiaries' or operators' faces, creating cognitive or emotional overload, ultimately reducing engagement. Nonprofit organizations should therefore tailor the presence of TO to the specific visual composition of each post, rather than adopting it as a blanket strategy. Looking ahead, this work lays the groundwork for more fine-grained investigations of TO design itself. Given the importance of TO as a driver of engagement, future research could examine how specific attributes, such as sentiment, configuration, font size, and placement within the image, shape audience response.

Exploring these design parameters would help translate our findings into detailed visual guidelines, enabling nonprofits to optimize the persuasive and aesthetic impact of their social media appeals.

At the same time, several broader avenues emerge. Future studies could assess whether these patterns hold across other social platforms with different visual norms (e.g., TikTok, Facebook, X) or in relation to temporal factors such as humanitarian crises or seasonal campaigns. Audience characteristics, including culture, age, and prior involvement with a cause, might also influence how viewers react to different text–image combinations. Methodologically, experiments and machine-learning approaches could complement our observational design to isolate causal mechanisms and identify optimal feature mixes at scale.

Finally, we acknowledge that our sample focuses on three large, English-speaking NGOs and that our measure of engagement (i.e., likes and comments) does not necessarily translate into concrete donations. Addressing these limitations will be essential for future research aimed at building a comprehensive understanding of how visual and textual elements drive not only online interaction but also tangible prosocial action.

References

Affonso, Felipe M. and Chris Janiszewski (2023), "Marketing by Design: The Influence of Perceptual Structure on Brand Performance," *Journal of Marketing*, 87 (53), 736–54.

Arnheim, Rudolf (1954), *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press.

Berger, J., A. Humphreys, S. Ludwig, W. W. Moe, O. Netzer, and D. A. Schweidel. 2020. "Uniting the Tribes: Using Text for Marketing Insight." *Journal of Marketing* 84 (1): 1–25. https://doi.org/10.1177/0022242919873106.

Chang, C. 2017. "Methodological Issues in Advertising Research: Current Status, Shifts, and Trends." *Journal of Advertising* 46 (1), 2-20.

Farace, S., Ordenes, F. V., Herhausen, D., Grewal, D., & Ruyter, K. de. (2025). Standing Out While Fitting In: Visual Design of Text Overlays in Social Media Communication. Journal of Marketing, 0(0). https://doi.org/10.1177/00222429251322773.

Hayes, A. (2018). Introduction to mediation, moderation, and conditional process analysis. A Regression-Based Approach. 2nd edition, New York: Guilford.

Henderson, John M., Phillip A. Weeks, Jr., and Andrew Hollingworth (1999), "The Effects of Semantic Consistency on Eye Movements During Complex Scene Viewing," *Journal of Experimental Psychology: Human Perception and Performance*, 25 (1), 210–28.

Huh, J., and E. C. Malthouse. 2020. "Advancing Computational Advertising: Conceptualization of the Field and Future Directions." *Journal of Advertising* 49 (4): 367–376. https://doi.org/10.1080/00913367.2020.1795759.

Kwon, J., Lin, H., Deng, L., Dellicompagni, T., & Kang, M. Y. (2022). Computerized emotional content analysis: empirical findings based on charity social media advertisements. *International Journal of Advertising*, 41(7), 1314–1337. https://doi.org/10.1080/02650487.2021.2012070.

Locher, P. J., Stappers, P. J., & Overbeeke, K. (1999). An empirical evaluation of the visual rightness theory of pictorial composition. *Acta psychologica*, 103(3), 261-280.

M+R. (2025). *2025 M+R Benchmarks Study*. Retrieved from https://mrbenchmarks.com.

Nikulina, O., van Riel, A. C. R., Lemmink, J. G. A. M., Grewal, D., & Wetzels, M. (2024). Narrate, Act, and Resonate to Tell a Visual Story: A Systematic Review of How Images Transport Viewers. *Journal of Advertising*, *53*(4), 605–625.

Ordenes, F. V., D. Grewal, S. Ludwig, K. D. Ruyter, D. Mahr, and M. Wetzels. 2019. "Cutting through Content Clutter: How Speech and Image Acts Drive Consumer Sharing of Social Media Brand Messages." *Journal of Consumer Research* 45 (5): 988–1012. https://doi.org/10.1093/jcr/ucy032.

Pieters, Rik and Michel Wedel (2004), "Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects," *Journal of Marketing*, 68 (2), 36–50.

Pieters, Rik, Michel Wedel, and Jie Zhang (2007), "Optimal Feature Advertising Design Under Competitive Clutter," *Management Science,* 53 (11), 1815–28.

Shi, Z., X. Liu, D. Lee, and K. Srinivasan. 2023. "How Do Fast-Fashion Copycats Affect the Popularity of Premium Brands? Evidence from Social Media." *Journal of Marketing Research* 60 (6), 1027-1051.

Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. Wiley.

Small, D. A., & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty*, *26*, 5-16.

Small, D. A., & Verrochi, N. M. (2009). The face of need: Facial emotion expression on charity advertisements. *Journal of Marketing Research*, 46(6), 777-787.

United Nations Development Programme. (2024). *10 tips to enhance social media presence for non-profits*. UNDP Blog.

Wooley, Kaitlin, and Marissa A. Sharif (2022), "The Psychology of Your Scrolling Addiction," *Harvard Business Review* (January 31), https://hbr.org/2022/01/the-psychology-of-your-scrolling-addiction.

Yazdani, E., Chakravarty, A., & Inman, J. (2025). Racial Inequity in Donation-Based Crowdfunding Platforms: The Role of Facial Emotional Expressiveness. *Journal of Marketing*, 89(4), 140-160. https://doi.org/10.1177/00222429241300320.

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197–206.