Visual Self-Presentation on Video-Based Social Media: A Pilot Study of Mental Illness Narratives on YouTube

Jiaying "Lizzy" Liu
jiayingliu@utexas.edu
School of Information, The University of Texas at Austin
USA

Yunlong Wang
wang_yunlong@ihpc.a-star.edu.sg
Institute of High Performance Computing (IHPC), A*STAR
Singapore

Allen Jue
mrallenjue@utexas.edu
The University of Texas at Austin
TX, USA

Yao Lyu
yaolyu@psu.edu
Pennsylvania State University
University Park, Pennsylvania, USA

Yiheng Su
Artificial Intelligence and Human-Centered Computing (AI&HCC) Lab
The University of Texas at Austin
Austin, Texas, USA
sam.su@utexas.edu

Shuo Niu
shniu@clarku.edu
Clark University
Worcester, MA, USA

Yan Zhang
yanz@utexas.edu
School of Information, The University of Texas at Austin
USA

Abstract

This pilot study explores the visual attributes of mental health self-presentation on social media, with a focus on YouTube as a key platform for video-mediated storytelling. We experimented with an LLM-Assisted Video Content Analysis Workflow to prepare for the quantitative investigation of visual attributes and audience engagement. By examining how visual techniques shape the portrayal of mental health, this study investigates their impact on audience perception and engagement. Additionally, it considers challenges such as visibility disparities and algorithmic mediation, which may marginalize certain narratives on video-sharing platforms. This research aims to contribute to HCI and Computational Social Science by deepening our understanding of visual self-presentation and its implications for individuals sharing mental health experiences online.

Keywords

Video-Sharing Platforms, Schizophrenia, Visual Disparity, Large Language Models

ACM Reference Format:

Jiaying "Lizzy" Liu, Yunlong Wang, Allen Jue, Yao Lyu, Yiheng Su, Shuo Niu, and Yan Zhang. 2025. Visual Self-Presentation on Video-Based Social Media: A Pilot Study of Mental Illness Narratives on YouTube. In *Proceedings of (CHI '25 Workshop on Envisioning the Future of Interactive Health).* ACM, New York, NY, USA, 4 pages.

CHI '25 Workshop on Envisioning the Future of Interactive Health, Yokohama, Japan

1 Background

Understanding illness narratives in online spaces, especially for socially stigmatized diseases, has been a central topic in Human-Computer Interaction (HCI) research. Scholars have found that technology-mediated disclosures, expressions, and discourses can contribute to both the empowerment and, unfortunately, sometimes the oppression and marginalization of certain illness experiences [6] and communities [20].

The emergence of video-based social media has revolutionized digital self-presentation, with platforms such as YouTube [13], TikTok [21], and Instagram [1] enabling rich, multimodal storytelling. The visual communication techniques employed on these platforms play a significant role in shaping both the creation and interpretation of video content, influencing audience perception [9] and engagement [12]. However, video-mediated communication also introduces challenges, including visibility disparities [17] and algorithmic mediation [11], which may systematically marginalize certain narratives. These challenges are particularly concerning for individuals with mental health conditions, who often rely on these platforms as vital spaces for sharing personal experiences and building community connections.

This study investigates the visual attributes of mental health self-presentation on social media, aiming to uncover how these attributes shape video-mediated communication and the portrayal of mental health. Focusing on YouTube is the primary research site given its prominence as a space where individuals with mental illness share their experiences through vlogs [10, 19]. The research seeks to address the following questions:

- RQ1: What visual self-presentation patterns characterize narratives of mental illness on YouTube?

- RQ2: How do specific visual elements correlate with audience engagement metrics and expressions of social support?

This study integrates Goffman's dramaturgical framework [7] to conceptualize video platforms as performative spaces where identity is enacted through multimodal elements [22], focusing on how creators strategically use visual aesthetics, performative techniques, and narrative strategies for self-presentation.

The research contributes to **social media studies**, **computational social science**, and **health behavior scholarship**: (1) By integrating visual culture analysis with digital media studies, it illuminates how visual affordances mediate contemporary mental health discourse; (2) the novel application of large language models (LLMs) to video content analysis advances computational methodologies while critically examining how platform architectures and algorithmic systems influence mental health narratives in digital spaces.

2 Research design

The research employs a mixed-method design, as illustrated in Figure 1, beginning with qualitative video analysis to explore nuanced patterns, followed by computational approaches to quantify and validate these findings at scale.

Step 1: Data Collection. The study analyzed schizophrenia vlogs, focusing on middle-length videos (4-20 minutes) for richer personal narratives. Using the YouTube Data API v3 with search queries such as "schizophrenia" and "schizophrenic," we collected 555 English-language videos from 2023. After filtering, 401 videos remained, with 200 randomly selected for analysis. The average duration was 9:23, and we sampled one frame per 30 seconds, yielding 3,800 frames.

Step 2: Theory-Informed Qualitative Coding. We employ iterative open and axial coding [5] to analyze visual representation techniques in videos, identifying key factors for regression analysis. Grounded in media theory, the framework examines video structures, frames, and transcripts, with a focus on visual elements—an underexplored area in HCI and social media research. Drawing on semiotic theory [3] and color theory, we decode visual signs (e.g., personal spaces, daily activities, interactions) and interpret color choices, where muted tones often signify depressive states and brighter hues appear in recovery narratives. This structured approach uncovers patterns in vloggers' self-representation and emotional disclosure.

Step 3: LLM-Assisted Video Annotation. Building on the qualitative coding framework, we employ a Multimodal LLM (MLLM) and text LLM to automate video feature annotation. This process follows our prior exploratory workflow published in [15, 16], incorporating iterative prompt design and human evaluation to ensure accuracy. In this study, we tested the workflow on 205 video frames using LLaVa-1.6[1] for four types of visual annotations.

Step 4: Regression Analysis. Using the LLM-assisted annotations, we perform regression analyses to investigate how visual and verbal features influence engagement metrics (e.g., views, likes) and comment-based support (e.g., informational support), evaluating the role of visual storytelling in shaping audience interaction.

[1]https://huggingface.co/llava-hf/llava-v1.6-mistral-7b-hf

3 Preliminary Findings

3.1 Qualitative Analysis of Videos

Iterative coding of videos revealed distinct visual features, as outlined in Table 1. Specifically, we identified two primary **video structures**: *talk-to-camera* and *in-the-moment*, which vloggers use to organize their content. Additionally, **frame-level** analysis uncovered patterns in how vloggers situate emotional disclosure through configurations of *vlogger presentation*, *stage setups*, and *stylistic shooting* choices.

3.1.1 Example-Stage. Overall, these visual features are closely tied to verbal narratives and appear to influence viewer engagement metrics and comment-based support. To illustrate this, I provide an example analysis of *stage setups*; for a comprehensive qualitative examination of the videos, please refer to [14].

The stage where vloggers present themselves is reflected in the disclosure of contexts and backgrounds of vlogs, representing a unique affordance of video disclosure compared to text posting on forums. *Space* reveals where vloggers record the videos, spanning a range of private spaces, such as homes and cars, as well as public areas like gyms and outdoor locations. Vloggers showcase *activities*, such as art-making, workouts, walks, and jogs. They may also involve *other people* such as family and friends onto the stage for various purposes.

We observed a spectrum of stage richness as depicted in the frames (Figure 2). In instances where the context is minimally disclosed, vloggers may position themselves in front of a plain background, such as a wall (a) or ceiling (b), or their faces may dominate the frame (c), leaving the surrounding context invisible. Conversely, some vloggers actively highlight their environment by showcasing detailed backgrounds and explicitly discussing their locations (d-f).

As suggested by Goffman, the "stage" as a performance front is intertwined with individuals' self-presentation. We observed that the details of vloggers' stages were closely tied to their verbal narratives. For example, one vlogger (d) filmed videos at his regular walking park while discussing the benefits of being close to nature, while another (e) sat in front of her bookshelf as she recommended books to viewers.

It seems that videos with a carefully crafted stage created opportunities for viewer interaction. Viewers expressed appreciation for vlogger (d)'s immersive walking experiences, with one commenting, "*Wow, what gorgeous scenery! Heavenly! My grandfather had peacocks... I enjoyed walking with you and listening to your talk.*" Vlogger (e) even received comments responding to details of her stage that she did not verbally mention, such as, "*I see you have an awesome collection of albums, can you do a video on your album collection?*"

3.2 Preliminary regression results

Based on the visual features identified in the qualitative study, we conducted a preliminary regression analysis to examine the relationship between video structure and engagement metrics. Due to limited computational resources, LLM-assisted annotation of the 3,800 frames is still ongoing. As an interim step, we present the proportions of video structures in Table 2. The results indicate that the proportion of *In-the-Moment* structures significantly increases

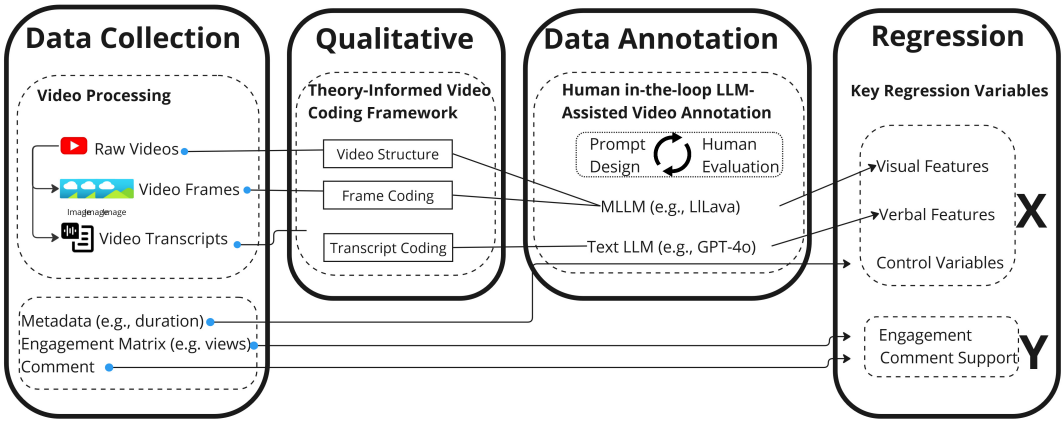

Figure 1: Research Design

Table 1: Summary of Qualitative Coding of Videos

	Features		
Video Structure	Talk-to-Camera		Vloggers address the camera directly with a static background, creating a confessional atmosphere.
	In-the-Moment		Vloggers visually portray their lives in a natural and vivid manner as they engage in various activities.
Frame-level	**Vlogger**	Anonymity	Whether vloggers reveal their facial identity.
		Demographics	Gender, race, age, etc.
		Identity	Religion, personal beliefs, and cultural background as conveyed in the vlogs.
	Stage	Space	The environment where vlogs take place, including background and setting.
		Activity	The vlogger's actions they engage in during the vlog.
		Other people	The presence of other people and their interactions.
	Style	Color	The use of color schemes and lighting in the frame to convey mood, tone, or emphasis.
		Aesthetics	The overall beauty and visual appeal of the images, related to spatial organization and visual hierarchy within the frame [18].
Transcript-level	Omitted for space concern in this manuscript		

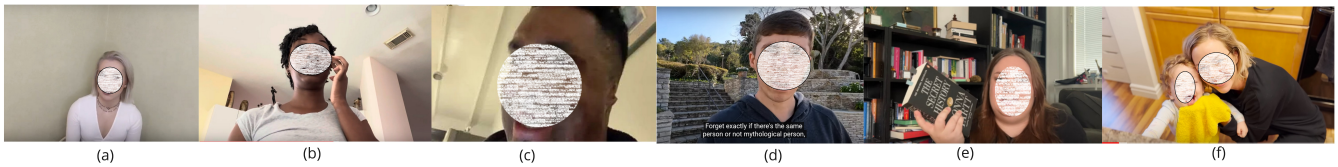

Figure 2: Stage

likes and comments, even after controlling for verbal features such as mentioned topics.

3.3 Summary of Preliminary findings

This work extends Goffman's dramaturgical framework to examine how mental health narratives are visually performed and received on video platforms. My analysis reveals how specific visual elements—from physical staging to aesthetic choices—mediate self-presentation of illness identity and audience engagement. For instance, strategic stage setup of personal spaces and immersive shooting (i.e., in-the-moment video structure) increased the perceived authenticity and viewer support.

4 Contribution

This study has the potential to advance two key areas in social media and online health community research:

A Visual Coding Framework for Video Self-Presentation. This work seeks to contribute a visual coding framework that builds on Hoffman's self-presentation theory [8] and visual analysis methods to explore the visual self-presentation patterns as reflected in user-generated videos. By analyzing aesthetic techniques, the framework highlights how visual elements shape mental health discourse beyond verbal narratives [22], while critically considering how algorithms may privilege certain visual styles [6]. This work

Table 2: Main effects in the regression model fitting of audience engagement and comment support.

		#Views (1)	#Likes (2)	#Comments (3)	Network Support (4)	Information Support (5)	Affirming (6)
Control	#Subscribers	**0.74** *** (0.06)	**0.81** *** (0.06)	**0.81** *** (0.06)	**0.09** ** (0.03)	0.01 (0.03)	0.03 (0.02)
	#Comments	—	—	—	**0.37** *** (0.03)	**0.45** *** (0.02)	**0.47** *** (0.02)
Visual Features	In-the-Moment	0.07 (0.06)	**0.19** ** (0.07)	**0.22** ** (0.07)	0.02 (0.03)	-0.01 (0.03)	**0.05** * (0.02)
Verbal Features	(Omitted in this manuscript for space concerns)						
Intercept	Intercept	**3.07** *** (0.06)	**3.21** *** (0.06)	**2.54** *** (0.07)	**0.48** *** (0.07)	**0.44** *** (0.06)	**0.61** *** (0.05)

Note: The numbers are the regression coefficients and the standard errors (in brackets). Statistically significant effects are in bold (* indicates p<.05, ** indicates p<.01, *** indicates p<.001.)

has the potential to bridge visual culture studies and digital communication, offering insights into how visual rhetoric influences both creator presentation and audience reception in social media contexts.

Computational Approaches to Video Analysis. This research aims to contributes a novel computational framework for analyzing video content, addressing gaps in prior work that has predominantly relied on textual data and basic image classification [4]. By examining visual rhetoric through framing structure, narrative context, and frame-level elements, the framework integrates LLM-assisted content analysis while maintaining a critical perspective on visibility dynamics [2] and algorithmic influence [11]. This approach has the potential to advance large-scale methods for analyzing visual rhetoric in video content, with implications for platform design, creator support, and algorithmic auditing.

References

[1] Nazanin Andalibi. 2017. Self-disclosure and Response Behaviors in Socially Stigmatized Contexts on Social Media: The Case of Miscarriage. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 248–253. doi:10.1145/3027063.3027137

[2] Kristen Barta and Nazanin Andalibi. 2024. Theorizing Self Visibility on Social Media: A Visibility Objects Lens. *ACM Transactions on Computer-Human Interaction* 31, 3 (June 2024), 1–28. doi:10.1145/3660337

[3] Roland Barthes. 1968. *Elements of Semiology*. Macmillan. Google-Books-ID: OVJhOA6iWxEC.

[4] Melissa Bica, Leysia Palen, and Chris Bopp. 2017. Visual Representations of Disaster. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1262–1276. doi:10.1145/2998181.2998212

[5] Juliet Corbin and Anselm Strauss. 2014. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications.

[6] Jessica L. Feuston and Anne Marie Piper. 2019. Everyday Experiences: Small Stories and Mental Illness on Instagram. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300495

[7] Erving Goffman. 2023. The presentation of self in everyday life. In *Social theory re-wired*. Routledge, 450–459. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003320609-59/presentation-self-everyday-life-erving-goffman

[8] Erving Goffman. 2023. The presentation of self in everyday life. In *Social theory re-wired*. Routledge, 450–459.

[9] Pengwei Hu, Chenhao Lin, Jiajia Li, Feng Tan, Xue Han, Xi Zhou, and Lun Hu. 2023. Making the Implicit Explicit: Depression Detection in Web across Posted Texts and Images. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 4807–4811. doi:10.1109/BIBM58861.2023.10385590 ISSN: 2156-1133.

[10] Jina Huh, Leslie S. Liu, Tina Neogi, Kori Inkpen, and Wanda Pratt. 2014. Health Vlogs as Social Support for Chronic Illness Management. *ACM Trans. Comput.-Hum. Interact.* 21, 4 (Aug. 2014), 23:1–23:31. doi:10.1145/2630067

[11] Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. 2021. Algorithmic Folk Theories and Identity: How TikTok Users Co-Produce Knowledge of Identity and Engage in Algorithmic Resistance. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 305:1–305:44. doi:10.1145/3476046

[12] Shuailin Li, Shiwei Wu, Tianjian Liu, Han Zhang, Qingyu Guo, and Zhenhui Peng. 2024. Understanding the Features of Text-Image Posts and Their Received Social Support in Online Grief Support Communities. *Proceedings of the International AAAI Conference on Web and Social Media* 18 (May 2024), 917–929. doi:10.1609/icwsm.v18i1.31362

[13] Jiaying Liu and Yan Zhang. 2024. Modeling Health Video Consumption Behaviors on Social Media: Activities, Challenges, and Characteristics. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (April 2024), 208:1–208:28. doi:10.1145/3653699

[14] Jiaying Liu and Yan Zhang. 2025. How do people with Schizophrenia Use YouTube Videos to Disclose Emotions? *preprint* (2025). https://13ea41b8-60a6-4134-a4c2-14ca423f3e44.filesusr.com/ugd/15d261_b057d715fc76454296b58dc75cf4c597.pdf

[15] Jiaying "Lizzy" Liu, Yiheng Su, and Praneel Seth. 2025. Can Large Language Models Grasp Concepts in Visual Content? A Case Study on YouTube Shorts about Depression. doi:10.1145/3706599.3719821 arXiv:2503.05109 [cs].

[16] Jiaying (Lizzy) Liu, Yunlong Wang, Yao Lyu, Yiheng Su, Shuo Niu, Xuhai "Orson" Xu, and Yan Zhang. 2024. Harnessing LLMs for Automated Video Content Analysis: An Exploratory Workflow of Short Videos on Depression. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing (CSCW Companion '24)*. Association for Computing Machinery, New York, NY, USA, 190–196. doi:10.1145/3678884.3681850

[17] Ryan McGrady, Kevin Zheng, Rebecca Curran, Jason Baumgartner, and Ethan Zuckerman. 2023. Dialing for Videos: A Random Sample of YouTube. *Journal of Quantitative Description: Digital Media* 3 (2023).

[18] Nikos Metallinos. 2013. *Television aesthetics: Perceptual, cognitive and compositional bases*. Routledge.

[19] Ashlee Milton, Leah Ajmani, Michael Ann DeVito, and Stevie Chancellor. 2023. "I See Me Here": Mental Health Content, Community, and Algorithmic Curation on TikTok. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3544548.3581489

[20] Sachin R. Pendse, Neha Kumar, and Munmun De Choudhury. 2023. Marginalization and the Construction of Mental Illness Narratives Online: Foregrounding Institutions in Technology-Mediated Care. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Oct. 2023), 346:1–346:30. doi:10.1145/3610195

[21] Anastasia Schaadhardt, Yue Fu, Cory Gennari Pratt, and Wanda Pratt. 2023. "Laughing so I don't cry": How TikTok users employ humor and compassion to connect around psychiatric hospitalization. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3544548.3581559

[22] Qian Wan and Zhicong Lu. 2024. Investigating VTubing as a Reconstruction of Streamer Self-Presentation: Identity, Performance, and Gender. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (April 2024), 1–22. doi:10.1145/3637357