

Implied vs. Expressed: Understanding the Impact of Social Media Posts on Users' Emotional Experience with LLM Predictions

ANONYMOUS AUTHOR(S)

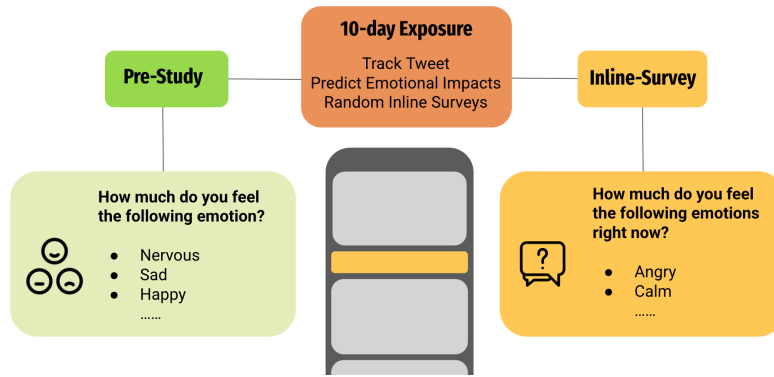


Fig. 1. The study design for the field experiment: The user fills out a pre-study survey before the experiment begins. During the 10-day exposure period, inline survey pops up as the user engages with the feed. All tweets viewed by the user are tracked and we obtain the emotional impact predictions from LLM.

Social media content has can significantly influence users' emotional states, yet understanding its precise emotional impact remains a critical challenge. LLMs offer potential in capturing implied emotional expressions, going beyond the detection of explicitly expressed emotions in traditional NLP methods. Our study leverages the capabilities of Large Language Models (LLMs) to predict the emotional impact of tweets and investigates how emotions in social media posts affect users' self-reported feelings. Preliminary findings reveal that aggregated LLM predictions of a tweet's emotional impact significantly predict users' self-reported emotions. The findings demonstrate that short-term exposure to social media content influences users' emotional experiences. This study presents a novel approach to examine emotional impacts by utilizing LLMs, which extend beyond traditional lexicon-based and semantic analysis methods.

Additional Key Words and Phrases: Emotion Evaluation and Emotion tracking, Emotion AI and Data, Large Language Models, Social Media, Sentiment Analysis

ACM Reference Format:

Anonymous Author(s). 2026. Implied vs. Expressed: Understanding the Impact of Social Media Posts on Users' Emotional Experience with LLM Predictions. 1, 1 (October 2026), 4 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

Manuscript submitted to ACM

1 Extended Abstract

Social media usage can significantly impact users' emotional experience [4]. While much psychology and cognitive science research has focused on the general behavioral usage of social media [8], whether and how social media content has an emotional impact on users remains a critical question. Understanding how users feel when spending time on the platform is critical to having a full understanding of their interactions and broader behavior—an angry user, for instance, may engage with the platform in fundamentally different ways, changing the dynamics of the online community and potentially compromising the quality of the discussions [5].

However, one critical challenge when investigating emotions in social media posts is that traditional NLP sentiment analysis tools and models primarily rely on direct language signals and focus on detecting users' **explicitly expressed emotions** [2, 7, 14]. Contrastingly, emotions can be conveyed through **implied methods** that do not involve explicit application of words associated with obvious affective connotations. People can interpret and perceive emotion on the sentence level, as opposed to words - the lexical building blocks[9]. The cultures, contexts, and common languages among certain groups also shape the way emotions are communicated [13], which are hardly captured by existing sentiment analysis tools. While explorations have been done in incorporating context in sentiment analysis, the context scope is normally represented restrictively as an extra lexicon that captures genre or concepts [11]. Other methods involve ML based techniques, but the trained classifiers are highly domain specific, as they require large scale labelled data in a particular context. **A significant gap remains between the emotions detected by sentiment analysis techniques and the emotional perceptions and impact to users' emotional experience.** This limitation is particularly relevant in social media, where posts frequently convey complex sentiments that are not easily captured by lexicon tools or domain-specific sentiment classifiers [1].

LLMs present potentials in closing this gap in sentiment analysis tasks. LLMs are trained on vast amounts of internet data encompassing diverse contexts so we believe they have a strong ability to predict the emotional impact of social media posts, even in complex scenarios where tweets convey implicit contextual sentiments or a mixture of multiple emotions. **Our project leverages the capabilities of Large Language Models (LLMs) to predict the emotions of social media posts. Using these predictions, we examine the impact of exposure to those posts on users' subsequent self-reported emotions.** We use data collected through a field study on X to investigate how the content users consume in their feeds directly impacts their emotions. The study utilizes a tweet dataset from *Social Media Algorithms Can Shape Affective Polarization* [12]. Users reported their emotions before the start of the analysis (pre-survey), and in inline surveys that were randomly populated into their feeds during a 10-day period (figure 1). We log the tweets that users browsed during the field experiment and employed the GPT-4 mini model and a one-shot prompt to predict how a tweet would make a user feel. Drawing on the arousal-valence theory [3, 10, 16] in psychology, we defined 15 affective states for the model to predict, incorporating definitions of each affect from psychology literature [6, 15] into the prompt. Given a tweet, the model was instructed to output its predicted rating for each emotion in JSON format. The prompt was refined through multiple iterations, with each version tested on a subset of 100 tweets and compared against the domain expert's annotations. We plan to further validate GPT's predictions with crowdsourced annotations.

There are 722,642 unique tweets in the initial dataset. After pre-processing (filtering out non-English tweets and those with fewer than six words), 258,915 unique tweets are viewed by the users. We obtained GPT emotion predictions for the tweets, which were viewed by users for more than one second. There are 1,233 participants in the analysis. A

total of 16,291 inline surveys were accounted for in the analysis. On average, each participant received 13 inline surveys over the course of the 10-day period.

A linear mixed-effects regression model was applied to account for variations in users' baseline emotion levels. To control for demographics and individual differences, we incorporated gender, age, and social status. Additionally, we included the corresponding pre-survey emotion to control users' emotional baselines before exposure.

$$\text{inline_survey_emotion} \sim (\text{gender} + \text{age} + \text{ladder}) + (\text{pre_survey_emotion}) + (\text{predicted_emotion})$$

The models predict the inline survey emotions based on the detected emotions from the tweets the users browsed. We examined the emotional impact of the tweet immediately before the inline survey, as well as the aggregated emotions of tweets viewed one hour prior to the inline survey. The emotional impacts of all tweets viewed by the user within the hour before the inline survey were aggregated using three methods: average, maximum, and exponential decay.

Preliminary findings shows that 1-hour aggregation of LLM annotations of tweet's emotional impact as significant predictor of user's self-reported emotions. Figure 2 presents the regression results. It indicates that the 1-hour average aggregation of GPT emotion predictions for tweets serves as a strong and significant predictor of users' self-reported emotions in the inline survey, particularly for anger, sadness, and excitement. However, GPT-predicted calm is not significantly correlated with user-reported calm in the inline survey across all four groups, likely due to the limited number of tweets associated with calm affect in the dataset viewed by users. The coefficients for 1-hour max, 1-hour exponential decay, and the single tweet viewed immediately before the survey remain relatively high for predicting anger. In contrast, the predictions for sadness and excitement appear to have a weaker effect on users' self-reported sadness and excitement.

	Anrgy		Excited		Sad		Calm	
Control/Indep. Variable	Coefficient	Stf. Err.	Coefficient	Stf. Err.	Coefficient	Stf. Err.	Coefficient	Stf. Err.
var_age	0.662	0.500	1.317	0.632	-1.025	0.533	2.401***	0.591
binary_female	-2.667**	1.014	-1.865	1.279	-0.702	1.093	-0.984	1.211
var_ladder	-0.445	0.291	3.178***	0.371	-1.445***	0.315	2.006***	0.346
var_pre_survey_angry	6.047***	0.501						
var_pre_survey_excited			11.316***	0.648				
var_pre_survey_sad					6.796***	0.545		
var_pre_survey_calm							9.347***	0.580
predicted_angry	3.705***	0.375						
predicted_excited			4.453***	0.516				
predicted_sad					3.707***	0.597		
predicted_calm							1.054	1.151

Fig. 2. The regression results for a 1-hour average aggregation show that angry, excited, and sad all have the LLM prediction as a strong and significant predictor of users' self-reported emotions in the inline survey. (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$)

The preliminary findings show that direct exposure to social media content measurably shapes user-reported emotional states. By analyzing the relationship between LLM predictions of a content's emotional impact and users' self-reported feelings, we demonstrate that even brief exposure to tweets can elicit detectable shifts in emotion.

The findings also highlight the potential of LLMs in measuring emotions in social media content. Connecting LLM predictions with user-reported emotions provides a valuable approach for studying emotional contagion at scale and offers insights for platform and algorithmic design. However, human validation of an LLM's annotation of a tweet's

emotional impact is necessary to assess the alignment between human perception and the LLM’s interpretation of emotions in a given text. Although LLMs present opportunities for understanding the complex nuances of implied emotions—going beyond traditional lexicon-based sentiment analysis—they also introduce risks related to biases, stereotypes, and potential harms in emotional prediction and impact. **Future studies should focus on a more comprehensive evaluation of LLMs’ ability to understand emotional expressions. Further research into LLMs’ capabilities—and the associated risks—in understanding, predicting, and annotating emotions could be highly beneficial for improving emotion prediction in social media and enhancing human-AI interactions.**

References

- [1] Murtadha Ahmed, Qun Chen, and Zhanhuai Li. 2020. Constructing domain-dependent sentiment dictionary for sentiment analysis. *Neural Computing and Applications* 32 (2020), 14719–14732.
- [2] Stefano Baccianella, Andrea Esuli, Fabrizio Sebastiani, et al. 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec*, Vol. 10. Valletta, 2200–2204.
- [3] Lisa Feldman Barrett and James A Russell. 1999. The structure of current affect: Controversies and emerging consensus. *Current directions in psychological science* 8, 1 (1999), 10–14.
- [4] Paul Best, Roger Manktelow, and Brian Taylor. 2014. Online communication, social media and adolescent wellbeing: A systematic narrative review. *Children and Youth Services Review* 41 (2014), 27–36.
- [5] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1217–1230.
- [6] Paul Ekman and Daniel Cordaro. 2011. What is meant by calling emotions basic. *Emotion review* 3, 4 (2011), 364–370.
- [7] Pollyanna Gonçalves, Matheus Araújo, Fabricio Benevenuto, and Meeyoung Cha. 2013. Comparing and combining sentiment analysis methods. In *Proceedings of the first ACM conference on Online social networks*. 27–38.
- [8] Ethan Kross, Philippe Verduyn, Gal Sheppes, Cory K Costello, John Jonides, and Oscar Ybarra. 2021. Social media and well-being: Pitfalls, progress, and next steps. *Trends in cognitive sciences* 25, 1 (2021), 55–66.
- [9] Vicky Tzuyin Lai, Roel M Willems, and Peter Hagoort. 2015. Feel between the lines: implied emotion in sentence comprehension. *Journal of Cognitive Neuroscience* 27, 8 (2015), 1528–1541.
- [10] Randy J Larsen and Edward Diener. 1992. Promises and problems with the circumplex model of emotion. (1992).
- [11] Aminu Muhammad, Nirmalie Wiratunga, and Robert Lothian. 2016. Contextual sentiment analysis for social media genres. *Knowledge-based systems* 108 (2016), 92–101.
- [12] Tiziano Piccardi, Martin Saveski, Chenyan Jia, Jeffrey T Hancock, Jeanne L Tsai, and Michael Bernstein. 2024. Social Media Algorithms Can Shape Affective Polarization via Exposure to Antidemocratic Attitudes and Partisan Animosity. *arXiv preprint arXiv:2411.14652* (2024).
- [13] Saba Safdar, Wolfgang Friedlmeier, David Matsumoto, Seung Hee Yoo, Catherine T Kwantes, Hisako Kakai, and Eri Shigemasu. 2009. Variations of emotional display rules within and across cultures: a comparison between Canada, USA, and Japan. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 41, 1 (2009), 1.
- [14] Mike Thelwall. 2016. The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In *Cyberemotions: Collective emotions in cyberspace*. Springer, 119–134.
- [15] Gary R VandenBos. 2007. *APA dictionary of psychology*. American Psychological Association.
- [16] David Watson and Auke Tellegen. 1985. Toward a consensual structure of mood. *Psychological bulletin* 98, 2 (1985), 219.