Reading the Digital Pulse: Context-Aware AI for Sensing Engagement Drivers in Health Forums

IEEE BSN 2025 Technical Paper

1st Pouria Rad
School of Computer & Cyber Sciences
Augusta University
Augusta, GA
PESLAMIRAD@augusta.edu

2nd Gianluca Zanella
School of Computer & Cyber Sciences
Augusta University
Augusta, GA
GZANELLA@augusta.edu

3rd Mohsen Jozani
Fowler College of Business
San Diego State University
San Diego, CA
MJOZANI@sdsu.edu

***Abstract*—Online Health Communities (OHCs) are vital for patient support, but the massive scale of user-generated content makes manual monitoring of user well-being and engagement drivers impossible. This study integrates Self-Determination Theory with a large-scale computational analysis of 1.03M questions and 6.0M replies, applying a fine-tuned DistilRoBERTa affect model (κ=0.74 vs. human) and Poisson regression (pseudo-R2=0.03) to quantify how emotional tone and informational intent predict engagement of over one million posts from the MedHelp platform. Our results show that, while informational contributions consistently predict higher engagement, the impact of emotional expression is highly context-dependent. For instance, expressions of sadness and anger, which have a minor effect on engagement overall, are strongly associated with increased community response in high-stakes forums like the Cancer community. These findings demonstrate the need for context-aware moderation tools and provide a scalable framework for designing more responsive and effective digital health interventions.**

***Index Terms*—digital health, emotion analytics, online health communities, decision support, adaptive intervention**

I. Introduction

Online Health Communities (OHCs) have emerged as critical digital ecosystems for patient-centered care, creating virtual spaces where millions of users exchange vital informational and socio-emotional support [1]–[3]. Yet, this supportive function is threatened by a fundamental challenge of scale. The sheer volume of user-generated text makes real-time monitoring by human moderators infeasible [4], meaning critical distress signals are often missed [5]. To better manage these platforms, we must first understand what drives user engagement. To do so, we ground our work in Self-Determination Theory (SDT), which posits that human behavior is driven by distinct extrinsic and intrinsic motivations [6].

In the context of OHCs, extrinsic motivations often manifest as informational contributions. Users may share knowledge by posting journal entries or answering questions across the platform to build a reputation, gain visibility, or out of a sense of reciprocity for help previously received [7]. These actions are instrumental in obtaining an outcome separate from the behavior itself. This leads to our first research question: To what extent do informational contribution behaviors predict community engagement?

Conversely, intrinsic motivations stem from the innate psychological needs for relatedness and positive emotional experiences [8]. The socio-emotional expression within a post is a direct signal of these needs. Past research suggests that the emotional tone of content is a powerful fuel for engagement, with positive, high-arousal emotions often increasing virality while deactivating negative emotions like sadness suppressing it [9], [10]. Understanding this dynamic is crucial for fostering supportive environments. Concurrently, multimodal models that fuse textual, affective, and interaction cues improve inference about informational content in online communities and generalize across sub-communities [11]. This brings us to our second research question: How does the emotional tone of a user's post relate to community engagement?

This study addresses the challenge of identifying the drivers of OHC engagement by answering these questions through large-scale computational analysis. Our proposed solution involves applying a fine-tuned, transformer-based AI model [12] to a dataset of over one million questions from MedHelp, allowing us to quantify the distinct impacts of informational motivations and emotional expressions on community response rates. This research offers three key contributions for practitioners. First, we provide a scalable and reproducible framework to diagnose the health of an online community. Second, we deliver actionable insights into how platform managers can foster specific user contributions. Finally, this work provides a clear pathway toward developing context-aware, data-driven tools to help moderators cultivate more vibrant and effective digital health ecosystems. The remainder of this paper details our methodology, presents our results, and discusses their broader implications.

II. Methodology

To answer our research questions, we employ a multi-step methodology. We begin by describing our large-scale dataset from a prominent OHC. Next, we detail our computational sensing approach, which uses a transformer-based AI model to extract features representing our theoretical constructs of emotional expression and informational motivation. Finally, we outline the specifics of our predictive model, a Poisson regression, chosen to best suit our count-based dependent variable.

A. Dataset and Sampling

Our data was collected from MedHelp (medhelp.org), a large online health community comprising, at the time of data collection, 270 sub-communities (forums), over 1.1 million user-posted questions, and approximately 6 million associated responses. After excluding non-English and non-health-related forums, our final dataset covered 137 distinct communities. For a focused analysis, we selected the four forums with the highest post counts: Cancer, Pain Management, Sexually Transmitted Infections (STI), and Women's Health. After removing posts with missing user data or unreadable text, our final analytic set comprises 735,966 questions (72% of the raw collection) which well exceeds the 2,436 observations required to achieve 80% power for a Poisson regression with an alpha of 0.05 [13], [14]. Descriptive statistics for the key variables used in our analysis are presented in Table I. Because the data are publicly available and no direct identifiers were collected, this work was deemed *not human-subjects research* under our University policy §4.2, consistent with 45 CFR 46.104(d)(4).

TABLE I
DESCRIPTIVE STATISTICS

Variable	**N**	**Mean**	**Std. Dev.**	**Min**	**Max**
Community Response	1.03M	5.39	4.85	0	24
Journals	0.98M	0.11	0.32	0	1
Communities	0.98M	2.60	4.48	0	151
PWRC	1.01M	241.92	1250.09	0	67498
Neutral	1.03M	0.20	0.22	0	0.97
Sadness	1.03M	0.20	0.25	0	0.99
Fear	1.03M	0.28	0.33	0	1.00
Anger	1.03M	0.05	0.12	0	1.00
Joy	1.03M	0.04	0.12	0	0.99
Friends	1.01M	0.19	7.49	0	3682
Gender (Female=1)	0.75M	0.75	0.43	0	1
Question Length	1.03M	124.15	130.80	1	3116
Notes	0.98M	14.63	96.22	0	3893
Photos	0.98M	0.86	3.15	0	16

B. Computational Sensing and Feature Engineering

Our primary dependent variable is Community Response Count, measured as the number of replies a question receives from other users (i.e., excluding self-replies). This serves as a direct proxy for community engagement and the two-pronged design—modeling socio-emotional signals alongside informational intent—follows evidence that jointly leveraging these channels improves informativeness detection in OHCs [11]. To operationalize the motivational constructs from SDT, we engineered two categories of features for our independent variables:

1. *Socio-Emotional Features (Intrinsic Motivation):* To quantify the emotional content of each question, we used a publicly available, fine-tuned DistilRoBERTa model [12]. This transformer-based model is specifically designed for multi-label emotion classification and is suitable for our context as it was trained on diverse text sources, including social media. For each question, we generated continuous scores (ranging from 0 to 1) for five of Ekman's fundamental emotions: Anger, Sadness, Joy, Fear, and Neutral [15].

2. *Informational Contribution Features (Extrinsic Motivation):* We created three variables to serve as proxies for a user's motivation to contribute informational content. Journals is a binary variable indicating if a user has posted "journal" entries, a feature on MedHelp for sharing information without soliciting answers. Communities is a count of the number of distinct forums a user is active in. Finally, Platform-Wide Response Count (PWRC) is a count of the total number of answers a questioner has provided to other users across the entire platform.

We also included several control variables, including the questioner's gender, number of friends, total uploaded Photos, received peer-to-peer messages (Notes), and the Question Length in words.

C. Predictive Modeling

Given that our dependent variable is a non-negative integer count, the most appropriate analytical approach is a Poisson regression [16]. This model is designed to predict the rate at which an event occurs based on a set of explanatory variables. Histograms of our dependent variable revealed a right-skewed distribution common to count data, with a high frequency of low response counts that rapidly trails off. A key assumption of the Poisson model is the equity of mean and variance. Our data did not exhibit significant overdispersion (mean=5.39, variance=23.52, not equal, but visually manageable) or problematic zero-inflation, allowing us to proceed with a standard Poisson model rather than a more complex variant like a zero-inflated or negative binomial model [17], [18].

To ensure our results are both statistically rigorous and practically meaningful, we report not only coefficients and p-values but also the Incident Rate Ratio (IRR) for each predictor. The IRR, which is the exponentiated coefficient, provides a standardized measure of effect size that quantifies the multiplicative change in the expected response count for a one-unit increase in a predictor [19]. This is crucial in large datasets where even trivial effects can become statistically significant, as it allows us to focus on the magnitude and practical importance of our findings [20].

III. RESULTS

We estimated Poisson regression models on the full dataset of over 735,000 questions and on subsamples for each of our four focal forums. Table II presents the full regression outputs.

In response to our first research question regarding informational contributions, the aggregate "All Forums" model shows that extrinsic motivational behaviors are strong predictors of engagement. Posting informational journal content is associated with a 22% increase in expected responses (IRR

TABLE II
POISSON REGRESSION RESULTS ACROSS FORUMS

DV: Community Response	All Forums	Cancer	Pain Management	STI	Women Health
Informational Level Variables (Extrinsic)					
Journals	0.198*** (0.00)	0.250*** (0.01)	0.165*** (0.01)	0.103*** (0.02)	0.073*** (0.00)
Communities	−0.002*** (0.00)	−0.006*** (0.00)	0.008*** (0.00)	0.011*** (0.00)	−0.001*** (0.00)
PWRC	0.000*** (0.00)	0.000*** (0.00)	0.000*** (0.00)	0.000 (0.00)	0.000*** (0.00)
Socio-Emotional Level Variables (Intrinsic)					
Neutral	0.031*** (0.00)	−0.018 (0.03)	−0.033 (0.03)	0.058*** (0.01)	0.040*** (0.00)
Sadness	−0.046*** (0.00)	0.243*** (0.02)	−0.229*** (0.03)	0.091*** (0.01)	−0.151*** (0.00)
Fear	−0.093*** (0.00)	−0.008 (0.02)	−0.068** (0.02)	0.053*** (0.01)	−0.136*** (0.00)
Anger	0.075*** (0.00)	0.200*** (0.04)	−0.080* (0.04)	0.049** (0.02)	0.133*** (0.01)
Joy	0.338*** (0.00)	0.643*** (0.03)	−0.126* (0.05)	0.036 (0.05)	0.224*** (0.01)
Control Variables					
Friends	−0.000 (0.00)	−0.003 (0.00)	−0.001 (0.00)	−0.000 (0.00)	−0.000 (0.00)
Gender	0.064*** (0.00)	0.204*** (0.01)	0.053*** (0.01)	−0.069*** (0.00)	0.294*** (0.01)
Question Length	0.000*** (0.00)	0.001*** (0.00)	0.001*** (0.00)	0.001*** (0.00)	0.000*** (0.00)
Notes	0.000*** (0.00)	0.001*** (0.00)	−0.000** (0.00)	−0.000 (0.00)	0.000*** (0.00)
Photos	0.013*** (0.00)	0.016*** (0.00)	0.007*** (0.00)	0.002 (0.00)	0.011*** (0.00)
Intercept	1.586*** (0.00)	1.166*** (0.02)	1.532*** (0.02)	1.308*** (0.01)	1.444*** (0.01)
Model Specifications					
Observations	735,966	14,476	9,426	46,204	301,972
LR chi2 (13)	132,663.68	6,864.49	2,103.60	2,713.41	45,969.68
Log Likelihood	-2,369,203.3	-42,710.48	-28,455.45	-127,033.43	-933,453.70
Pseudo R2	0.0272	0.0744	0.0356	0.0106	0.0240

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Standard errors in parentheses.

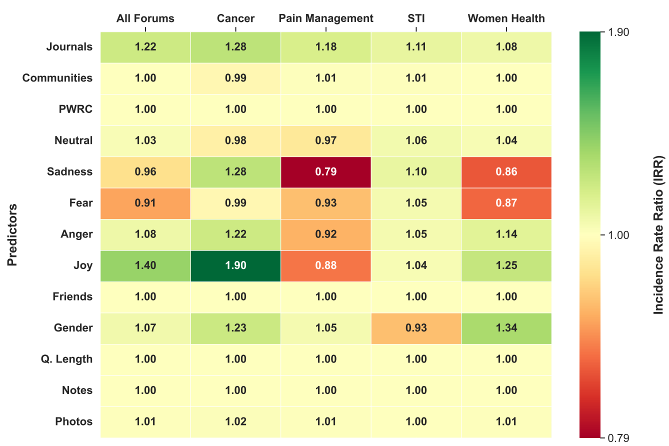

Fig. 1. Incident Rate Ratios (IRR) of Predictors on Community Response Count. Green indicates a positive effect (IRR > 1), red indicates a negative effect (IRR < 1).

= 1.22, p $<$ 0.01). Likewise, a user's history of answering other questions (PWRC) is positively, though very slightly, associated with engagement. In contrast, being active in many communities has a negligible negative effect overall, suggesting that breadth of participation does not necessarily translate to deeper engagement on any single question.

Regarding our second research question on socio-emotional expression, the emotional tone of a post is a powerful and nuanced predictor. Across all forums, posts expressing joy receive a remarkable 40% more responses on average (IRR = 1.40, p $<$ 0.01), while expressions of sadness and fear are associated with a 4% and 9% reduction in responses, respectively. Strikingly, posts with a high anger score receive 8% more responses, running counter to the notion that all negative emotions suppress engagement.

However, these aggregate findings mask significant and managerially important variations across different health contexts. The practical impact of these differences is best visualized by the heatmap of Incident Rate Ratios (IRR) in Figure 1. An IRR above 1.0 (green) indicates a positive effect on response count, while a value below 1.0 (red) indicates a negative effect, with color intensity showing the magnitude. For example, while joy is a strong positive predictor overall, it is associated with a 12% decrease in engagement in the Pain Management forum (IRR = 0.88). Even more dramatically, expressions of sadness and anger, which suppress or slightly boost engagement overall, are associated with a 28% and 22% increase in responses in the Cancer forum, respectively. This suggests that in communities centered on severe, high-stakes conditions, expressions of distress may function as a strong call for support, fundamentally altering engagement dynamics.

IV. Discussion and Implications

This study leveraged computational sensing to explore the motivational drivers of engagement in OHCs, offering insights for both theory and practice. Our findings demonstrate that while informational contributions are a consistent, positive driver of engagement across communities, the impact of emotional expression is highly context-dependent, providing a nuanced view of Self-Determination Theory in digital health spaces.

A. Practical Implications for Digital Health

Our results offer several actionable implications for the design and management of OHCs. First, the consistent positive effect of "Journal" entries suggests that platforms should actively encourage and feature informational content. By creating dedicated spaces for users to share knowledge without the expectation of an immediate answer, platforms can foster a repository of valuable content that stimulates reciprocal engagement throughout the community. A complementary pathway is agentic, context-aware coaching that integrates continuous tracking, personalized Q&A/education, and social gamification, an approach consistent with recent AI health-coach designs [21].

Second, the striking variance in emotional impact across forums is our most critical finding for practice. A "one-size-fits-all" approach to community management is suboptimal. In the Cancer forum, for instance, sadness and anger are powerful signals for support; moderators in this context could be trained to prioritize posts high in these emotions for intervention. Conversely, in the Pain Management forum, expressions of joy are negatively associated with engagement, perhaps because they are seen as off-topic or unrelatable. The results inform practitioners by showing how crucial it is to foster the socio-emotional factor to maximize engagement. While community contribution is often rewarded with features like a 'Best Answer' star, our work suggests the need for new metrics and dashboard tools that track and encourage context-specific emotional support, which can directly improve intrinsic motivations.

B. Limitations and Future Work

This study, while extensive, has limitations that open avenues for future research. Our cross-sectional analysis identifies strong associations but cannot establish causality. Future work could employ longitudinal methods to track how changes in a user's emotional expression and contribution patterns over time influence their engagement.

The most promising direction for future work lies in building and testing adaptive interventions based on our findings. For example, one could develop a system that detects a post high in sadness in the Cancer forum and automatically alerts a "super-user" or community champion to respond. Similarly, an AI Agent could be designed to offer supportive resources to users whose posts in the Pain Management forum express high levels of joy but receive few responses, recognizing this as a potential moment of disconnection. By moving from passive sensing to active intervention, we can fully realize the potential of computational methods to build more responsive, supportive, and effective online health communities.

References

[1] M. Van der Eijk, M. J. Faber, J. W. Aarts, J. A. Kremer, M. Munneke, and B. R. Bloem, "Using online health communities to deliver patient-centered care to people with chronic conditions," *Journal of medical Internet research*, vol. 15, no. 6, p. e115, 2013.

[2] G. Eysenbach, J. Powell, M. Englesakis, C. Rizo, and A. Stern, "Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions," *Bmj*, vol. 328, no. 7449, p. 1166, 2004.

[3] X. Zhang, S. Gao, Y. Cheng, and F. Meng, "Encouraging physicians' continuous knowledge-sharing in online health communities: A motivational perspective," *Frontiers in Public Health*, vol. 10, p. 1036366, 2022.

[4] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International journal of information management*, vol. 35, no. 2, pp. 137–144, 2015.

[5] J. Zhou, Q. Zhang, S. Zhou, X. Li, and X. M. Zhang, "Unintended emotional effects of online health communities: A text mining-supported empirical study." *Mis Quarterly*, vol. 47, no. 1, 2023.

[6] E. L. Deci and R. M. Ryan, *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media, 2013.

[7] M. M. Wasko and S. Faraj, "Why should i share? examining social capital and knowledge contribution in electronic networks of practice," *MIS quarterly*, pp. 35–57, 2005.

[8] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." *American psychologist*, vol. 55, no. 1, p. 68, 2000.

[9] M. Schreiner, T. Fischer, and R. Riedl, "Impact of content characteristics and emotion on behavioral engagement in social media: literature review and research agenda," *Electronic Commerce Research*, vol. 21, pp. 329–345, 2021.

[10] X. Wang and E. W. Lee, "Negative emotions shape the diffusion of cancer tweets: toward an integrated social network–text analytics approach," *Internet Research*, vol. 31, no. 2, pp. 401–418, 2021.

[11] M. Jozani, J. A. Williams, A. Aleroud, and S. Bhagat, "Emotional and informational dynamics in question-response pairs in online health communities: A multimodal deep learning approach," *Information Systems Frontiers*, pp. 1–25, 2025.

[12] J. Hartmann, "Emotion english distilroberta-base," https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/, 2022.

[13] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, "Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses," *Behavior research methods*, vol. 41, no. 4, pp. 1149–1160, 2009.

[14] J. Cohen, "Things i have learned (so far)." in *Annual Convention of the American Psychological Association, 98th, Aug, 1990, Boston, MA, US; Presented at the aforementioned conference.* American Psychological Association, 1992.

[15] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[16] G. King, "Variance specification in event count models: From restrictive assumptions to a generalized estimator," *American Journal of Political Science*, pp. 762–784, 1989.

[17] D. Lambert, "Zero-inflated poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, no. 1, pp. 1–14, 1992.

[18] J. M. Hilbe, *Negative binomial regression*. Cambridge University Press, 2011.

[19] S. Coxe, "Effect size measures for nonlinear count regression models," in *American Statistical Association Conference on Statistical Practice*, 2018, pp. 15–17.

[20] G. M. Sullivan and R. Feinn, "Using effect size—or why the p value is not enough," *Journal of graduate medical education*, vol. 4, no. 3, pp. 279–282, 2012.

[21] M. Jozani, N. Ebadi, R. Khan, and D. Zoughbie, "Chatgpc: A novel interactive ai health coach for obesity and diabetes prevention," 2025.