

Original Paper

# Forging Online Community Among People in Recovery From Substance Use: Natural Language Processing and Deep-Learning Analysis of The Phoenix App User-Generated Data

Danny Valdez<sup>1</sup>, PhD; Katie M Heinrich<sup>2</sup>, PhD; Beth Collinson<sup>2</sup>, PhD; Aspen Streetman<sup>3</sup>, PhD; Zach Sloan<sup>4</sup>, MBA

<sup>1</sup>Department of Applied Health Science, Indiana University, Bloomington, IN, United States

<sup>2</sup>Department of Research and Evaluation, The Phoenix, Boston, MA, United States

<sup>3</sup>Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, United States

<sup>4</sup>The Phoenix, Boston, MA, United States

## Corresponding Author:

Katie M Heinrich, PhD

Department of Research and Evaluation

The Phoenix

54 Newmarket Sq

Boston, MA 02118

United States

Phone: 1 8084579525

Email: [kmhphd@gmail.com](mailto:kmhphd@gmail.com)

## Abstract

**Background:** Mobile apps are powerful tools for promoting and sustaining healthy behaviors, including supporting diverse recovery pathways from substance use, including alcohol use disorder. Indeed, prior research strongly supports the notion that social connection through mobile platforms, supplemented by an in-person interaction, is vital in helping individuals strengthen their recovery and improve overall well-being and mental health. However, research into the digital footprints of mobile app users, as a strategy to assess app usage experiences in a recovery context, is lacking.

**Objective:** This study utilizes a dataset from The Phoenix app, a social media platform specifically designed for individuals impacted by substance use, including those in or seeking recovery, to identify core uses of the app, including how it is leveraged by members from a thematic and emotional valence context.

**Methods:** We applied natural language processing and deep learning methods to analyze a random sample of 19,685 posts. Analyses included the Bidirectional Encoder Representation from Transformers topic modeling tool to generate themes and a Valence Aware Dictionary and Sentiment Reasoner sentiment analysis to approximate emotional tone and mood from posts ranging from highly negative (−0.99) to highly positive (0.99).

**Results:** After removing duplicate and nonsensical posts, we retained a final sample size of 17,617 posts. Bidirectional Encoder Representation from Transformers topic modeling tool identified 10 topics (coherence score=0.48) within 2 overarching themes: (1) those related to engaging app members through in-person and online interactions (7 topics) and (2) as a forum to discuss more serious topics pertaining to substance use and mental health recovery (3 topics). Overall, the topics revealed a distinct and recurring theme of community support. Valence Aware Dictionary and Sentiment Reasoner sentiment analysis was 0.44 (SD 0.42), indicating highly positive posts, with only 429 (2.4%) being highly negative.

**Conclusions:** The study findings broadly show positive uses of The Phoenix app as a tool for social connections and community among people in recovery from substance use. With the high positive sentiment of posts, the app was distinct from other social media platforms (eg, X, Reddit, Facebook), which often feature a mix of highly positive and highly negative posts. Additional research is needed to confirm these results using a larger dataset and with comparative analysis of other recovery forums to contribute to the understanding of social media's role and function in changing health-related behaviors.

*JMIR Mhealth Uhealth* 2025;13:e68438; doi: [10.2196/68438](https://doi.org/10.2196/68438)

**Keywords:** community; mental health; mHealth; mobile health; social support; substance use

## Introduction

### Background

Recovery is a behavioral health process among individuals affected by substance use and substance use disorders (SUDs) [1]. In 2012, the Substance Abuse and Mental Health Services Administration developed a new working definition of recovery, defining the process as “a...change through which individuals improve their health and wellness, live self-directed lives, and strive to reach their full potential” [2]. This definition shifts away from the traditional emphasis on abstinence along with symptom mitigation and reframes recovery as a holistic, multidimensional, and humanistic process [3]. By emphasizing overall well-being, the Substance Abuse and Mental Health Services Administration broadens the focus to include improvement in multiple life domains for individuals affected by substance use or other similar challenges. This shift underscores the importance of individualized and comprehensive recovery pathways that are accessible and not time-bound and foster supportive environments to promote recovery outcomes [4,5].

Emphasizing recovery is critical given the evolving US substance use crisis, which affects individuals across age groups and demographics [6]. As many as one-third of US adults reported an SUD in the past year, and the number of people facing substance-related challenges is growing substantially [7,8]. Yet, in the United States, most adults with SUD do not receive treatment [6] due to well-documented barriers to care that include cost, geographic inaccessibility, perceived stigma, and the availability of co-occurring disorder treatment [9–11]. Novel strategies that address this crisis with approaches that promote engagement in recovery by meeting people where they are in ways that are both accessible and intuitive are needed.

### Mobile Apps as Accessible Recovery-Promoting Tools

Growing evidence suggests that social media forums and smartphone apps may offer practical and cost-effective strategies for behavior change, which can significantly decrease substance use [12–14]. Some of the most frequently deployed behavior change strategies used on these forums and apps include self-monitoring, personalization, reminders [14], and online peer coaching [15]. These app features enable users to tailor the app’s functionality to meet their individual needs, resulting in personalized recovery plans, increased app use, prolonged engagement in recovery, and reductions in substance use [13,14,16,17].

Social media platforms, and their respective smartphone apps, also help facilitate social support for substance use and mental health recovery [18–21]. Previous research shows, for example, that online recovery forums on Reddit can provide a sense of belonging and provide practical advice for individuals with SUD through interactions with others who share lived experiences and peer coaches trained to discuss substance use virtually [19,22]. Additionally, hashtags and content related to recovery on platforms such as X and

TikTok can raise awareness; reduce stigma; and connect individuals with resources, professional counseling, and other support networks [18,21]. Indeed, these virtual spaces allow people to share their experiences and seek guidance from others who have faced similar challenges, which can potentially support the engagement and connection necessary for promoting behavior change [23].

Other digital smartphone apps, which include closed social media–like platforms designed specifically for people in recovery, provide accessible, highly specific, and flexible opportunities for individuals to engage with recovery networks, often bypassing barriers known to impede care such as stigma [24]. Research likewise suggests that individuals are open to the idea of using social media or similar apps for recovery support, sometimes citing convenience and semianonymity in these spaces [25,26].

Recovery-specific digital platforms or apps—that is, InTheRooms and The Phoenix—have emerged as online resources beyond conventional social media designed to foster a community among people with lived experiences of substance use [23,25,27,28]. Although more research is needed on the long-term impacts of these apps, early indications suggest that they offer recovery-specific tools that provide emotional and practical support, paralleling the benefits of in-person interactions [29].

### Purpose

Individuals wanting to change behaviors may seek online and offline resources to initiate or sustain recovery from substance use and related challenges [30–32]. This study examines The Phoenix: A Sober Community, a nonprofit organization with a nationwide reach that developed a free mobile app (hereafter referred to as The Phoenix app, now known as NewForm). The app was designed to facilitate access to social support across geographical and temporal boundaries. It has since evolved into a robust platform that allows members and volunteers to connect based on shared interests or locations (eg, sober parents and guardians, veterans, arts, and crafts). It includes features such as live-streamed and in-person events, recovery journey tracking, volunteer coordination tools, and peer-to-peer communication.

As a social media app with a closed environment (ie, tailored for a specific purpose), understanding the app’s thematic and emotional digital ecosystem may offer additional insights into the role of digitally delivered peer- and volunteer-mediated interactions in promoting a safe and inclusive recovery environment. This study aims to identify the thematic and emotional climate of The Phoenix app using the in-the-moment interactions from anonymized app member data. Our study is guided by 3 research questions (RQs):

1. Using natural language processing and deep learning, what themes emerge from a corpus of posts originating from The Phoenix app?
2. What do compound sentiment analysis scores implicate regarding the personal and collective experiences of The Phoenix app members?

3. How can data mining strategies inform the future adaptation of The Phoenix app and member experiences?

The findings from this study will contribute to the growing body of evidence calling for directed and curated app experiences to promote behavior change in the context of substance use. In particular, the methods, or pipeline, undertaken in this study will also illustrate their effectiveness in highlighting app functionality along several critical domains, including language (ie, words and phrases used on mobile apps), emotions (ie, valence patterns used in mobile apps), and behaviors (ie, identifying observations in which recovery outcomes are salient). Collectively, this pipeline—which uses user-generated content provided by The Phoenix—will further document successful strategies necessary to sustain healthy digital environments for substance use recovery.

## Methods

### *Data and Sample Allocation*

#### Data

Data for this study were provided to the research team directly by The Phoenix and originated from The Phoenix app. The aim of our study was to test an unsupervised natural language processing (NLP) pipeline as part of a larger research project aimed at improving the functionality and usability of the app. These analyses are also intended to create a scaffold for building more refined machine learning and artificial intelligence algorithms trained to understand recovery-related language and needs.

By downloading and using the app, members agree to The Phoenix privacy policy, including consent to their anonymized data from the app for research purposes at The Phoenix. To uphold anonymity, posts derived from The Phoenix app are first deidentified by The Phoenix. During the anonymization phase, all personally identifiable information is removed, including a member's name, location, and other personally identifiable information. Once data have been anonymized and deidentified, they are assigned a "Text ID" to facilitate easier comparisons and faster identification by the core data analysis and research teams. Once data have been completely deidentified, they are stored in a single .csv file and distributed to additional members of the research team, including a computational data analyst and behavioral research scientists.

#### Sample

We were broadly interested in the function and utility of The Phoenix app to build community and promote recovery behavior. Given our outlined research questions and project aims, we sought to analyze a robust collection of randomly allocated individual posts originating from The Phoenix app. The dataset included a random collection of 19,685 app entries posted within a 6-month period from October 30, 2023, to April 3, 2024. This sample included 3 endpoints:

text ID, anonymized author ID, and post text. After removing duplicates, nonsensical data, and posts comprising random strings, we retained a final sample size of 17,617 posts, which we further processed through our computational pipeline.

### *Ethical Considerations*

Because this study involves the analysis of deidentified and secondary data, it was deemed as nonhuman subjects research and did not require further review by the institutional review board at Indiana University Bloomington (#24091). Given the sensitive nature of the data, preliminary findings from this study were shared with The Phoenix app members through open-access listening sessions. Any excerpts published in this report have been abridged slightly to uphold app member privacy by removing potentially identifiable information, graphic depictions of substance use, and other personal details.

## Analyses

### Overview

We used unsupervised NLP and deep learning methods to compositely analyze all posts included for analysis. The analyses used in this study include the Bidirectional Encoder Representation from Transformers topic modeling tool (BERTopic) for theme generation and a Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment analysis as an approximation for mood and well-being. Both computational analyses are briefly explained in the next sections.

### BERTopic Tool

BERTopic is a state-of-the-art deep learning topic modeling tool that approximates themes from a collection of similar texts or corpora [33]. Unlike conventional topic modeling analyses, which include latent Dirichlet allocation and latent semantic analysis, BERTopic uses embeddings calculated from a neural network architecture that compares an input sequence against a pretrained large language model [34]. Embeddings calculated by BERTopic are then used to convert the input sequence into continuous vectors, the dimensions of these vectors are reduced, and text-clustering algorithms help extract the most salient themes or ideas in the data. Although topics derived from BERTopic are, in most cases, conceptually clear, human interpretation can help address potential gaps and misclassifications. As such, a manual review of computer-derived topics is recommended to both ascribe meaning to the topics and attain agreement on the meaning among research personnel. For more information on BERTopic and its application in social media user-generated data, refer to Edinger et al and Valdez et al [35,36].

### VADER Analysis

VADER is a rule-based sentiment analysis lexicon commonly applied to social media text data to measure its valence or affect [37,38]. VADER scores can range from highly negative (−0.99), as in "everything is terrible right now," to highly positive (0.99), as in "this is the best day of my entire

life!” Despite being an older sentiment analysis tool, VADER offers several advantages, including its low computational power requirements and wide range of possible scores, which may reflect a broader range of emotions. These scores can be applied in various general linear modeling methodologies, such as regression and ANOVA. For further details on VADER, the VADER lexicon, and the validation of VADER, see the following [39].

## Text2Emotion

While VADER is widely regarded as an effective tool for sentiment analysis with social media data, we also ran a Text2Emotion analysis—a similar sentiment analysis tool that captures 5 discrete emotion categories: (1) Happy, (2) Sad, (3) Angry, (4) Surprise, and (5) Fear—for nuance. Scores along these 5 emotions are normalized values from 0 to 1, summing to 1 for each text entry, and reflect the intensity of each measure emotion. VADER and Text2Emotion scores can be compared to substantiate each analysis.

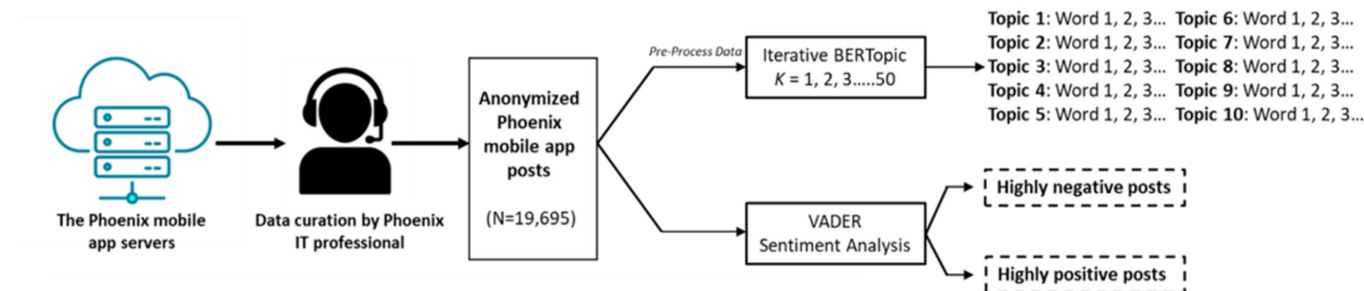
## Procedure

The steps followed in this study are shown in Figure 1. We first preprocessed the data to ensure standard and consistent entries. For our BERTopic analysis, preprocessing steps included removing parts of speech with no clear semantic meaning, such as (1) articles, (2) prepositions, (3) punctuations, (4) abbreviations, (5) any special characters, and (6) numbers. After preprocessing, we ran iterative BERTopic analysis, testing a range of possible topic solutions ( $k=5\ldots 50$ )

to find an optimal topic number. We calculated a coherence score with each iteration as an indicator of model fit, where coherence scores reflect the human interpretability of a fixed set of topics. For each iteration, we plotted each coherence score, with the highest coherence score indicating the optimal solution. Upon running 50 topic iterations, the topic solution with the highest coherence score was 10 (coherence=0.48). Once we identified a topic solution, we applied UMAP and HDBSCAN to reduce dimensions and extract topics with key words (eg, topic 1: Word a, Word b, Word c...Word d). We then used a simple sorting function to match each app entry to a topic based on the overlap of words used in the entry to topic key words. After sorting our topics, the members of the research team met briefly to agree on an interpretation of each topic.

For the VADER analysis, we fed minimally processed text through the VADER lexicon, retaining stop words, punctuations, numbers, and special symbols and characters. Because VADER requires exact-case matching to return an accurate score, that is, VADER can only provide a score if the spelling is correct, we used TextBlob and a loop function to replace apparent typos such that “angryyyy” is automatically corrected to its apparently correct spelling of “angry.” Once each entry received a VADER score and Text2Emotion profile, we created subdatasets consisting of highly negative entries (−0.99 to −0.50) and highly positive entries (0.50 to 0.99), which were reviewed by the research team to ascertain themes.

**Figure 1.** Computational pipeline from data allocation to topic modeling and sentiment analysis. BERTopic: Bidirectional Encoder Representation from Transformers topic modeling tool; VADER: Valence Aware Dictionary and Sentiment Reasoner.



## Results

### Overview

Broadly, our topic modeling analyses found various topics specific to promoting social connection; our sentiment analysis revealed how people utilize the app to seek or provide social support to those struggling with their recovery journeys. We present the findings from our analysis below.

### RQ1: Using Natural Language Processing and Deep Learning, What Themes

### Emergence From a Corpus of Posts Originating From The Phoenix App?

From the iterative BERTopic analysis, it was determined that 10 topics represented the optimal solution with a coherence score of 0.48. We then extracted top terms per topic and their representation as a whole number and percentage of the larger corpus (Table 1), and based on topic keywords, we created a custom name for each of the 10 topics. Topics extracted from our BERTopic analysis revealed what appeared to be 2 overarching themes: (1) a group of topics pertaining to engaging members through in-person and online interactions (7 topics) and (2) as a forum to discuss more serious topics pertaining to substance use and mental health, including struggles with sobriety, relapse, and strategies to promote substance use reduction or abstinence and improve



mental health (3 topics). These overarching themes are again observed in Figure 2, which visualizes an intertopic correlation as an intertopic distance map. Dense and closely knit circles (ie, topics) indicate high topic similarity, while disparate topics represent dissimilarity. The topics in the upper left-hand quadrant of the intertopic distance map largely pertain to app dialog about promoting in-person and online events, which ranged from hikes (eg, “Don’t forget about The Phoenix hike tonight at 8 pm!”), workouts (eg,

“See you tonight at...the gym for a great workout session”), online and offline poetry sharing (eg, “I am pasting my poem here; let me know what you think”), and other events deemed part of community building among app members. In contrast, the topics in the bottom-right quadrant, while sharing some overlap with the topics in the upper left quadrant, largely pertained to sharing personal experiences; narratives; or insights into substance use, recovery, and substance relapse.

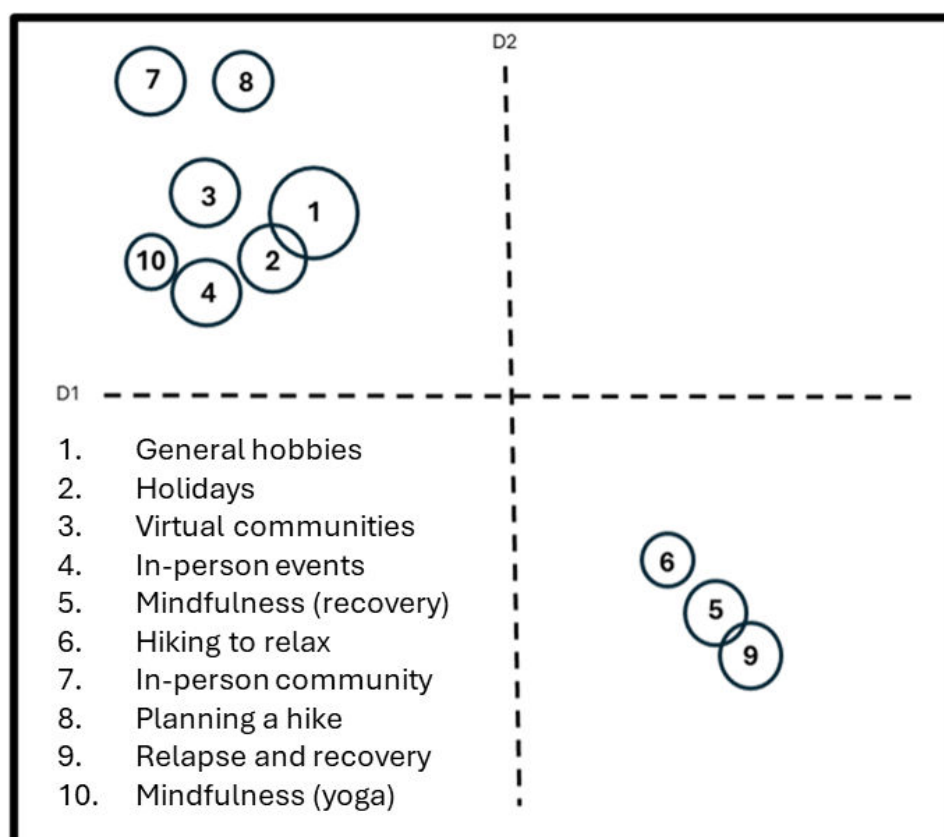
**Table 1.** BERTopic<sup>a</sup> summary table with 10 topics, top terms per topic, post count (n=17,617), and percentage of the larger corpus<sup>b</sup>.

ID	Custom name	Top terms	Post count (%)
1	General hobbies	favorite - movie - house - film - kitty - best - loved - baby - drew - lil	3417 (19.4)
2	Holidays	holiday - thanksgiving - weekend - Sunday - morning - celebrate - Christmas	1925 (10.9)
3	Virtual community	meet - connect - chat - joined - meeting - event - online - poem - site - join	1826 (10.4)
4	In-person events	festival - Saturday - Wednesday - weekend - fest - event - Thursday - tonight - attend	1785 (10.1)
5	Mindfulness	meditation - prayer - recovery - soul - felt - self - life - care - gratitude - change	1621 (9.2)
6	Hiking to relax	hike - stroll - walk - park - hiking - canyon - trail - leashed - sobriety - trip	1543 (8.8)
7	In-person community	sober - rehab - sobriety - meet - community - recovery - meeting - join - phoenix - hi	1491 (8.5)
8	Planning a hike	hike - hiking - trail - location - nj - area - az - hiker - valley - boulder	1447 (8.2)
9	Relapse	relapse - rehab - relapsed - substance use - addict - sobriety - sober - recovery - drug - stay	1365 (7.8)
10	Mindfulness	sobriety - recovery - walking - fitness - walked - yoga - volunteer - daily - exercise	1197 (6.8)

<sup>a</sup>BERTopic: Bidirectional Encoder Representation from Transformers topic modeling tool.

<sup>b</sup>Per topic, we extracted the 10 most salient terms. These terms were used to match app entries to a corresponding topic. The most prominent topic in our data, “general hobbies,” comprised 19.4% (n=3417) of all entries, while the least prominent topic was “mindfulness,” comprising 6.8% (n=1197) of all entries.

**Figure 2.** This visualization is an intertopic distance map highlighting the relative similarity between the topics. Around 7 topics appear in the upper left-hand quadrant, and 3 topics appear in the bottom-right quadrant. This suggests 2 overarching themes within the data.



Upon reviewing topic keywords and intertopic similarities, we then used a sorting function to sort posts into one of the 10 possible topics based on keyword matching. From this sorting, we were able to identify a series of optimal

quotations, which added further context and nuance to the topics explained above (Table 2). Table 2 shows the actual conversations unfolding with greater nuance from the dataset.

**Table 2.** Representative excerpts from the dataset per topic, parsed by topic ID, topic name, and key excerpts<sup>a</sup>.

Topic ID	Topic name	Key excerpts
1	General hobbies	<ol style="list-style-type: none"> <li>1. <i>Hey beautiful people. What an amazing way to connect. I need new friends in the area. My hobbies: art, music, hair, meetings (lol!), beach!</i></li> <li>2. <i>Its honestly been so long since I have done anything social, so I am just figuring out what my hobbies and things I like to do are!</i></li> <li>3. <i>Hi friends! Looking to build a strong sober friend group. I have lots of hobbies, passions and ideas for groups. Let's be friends!</i></li> </ol>
2	Holidays	<ol style="list-style-type: none"> <li>1. <i>Merry Christmas and Happy Holidays from our family to yours!</i></li> <li>2. <i>9am X-mas Eve X-fit crushed it this morning! Happy holidays to all! Come out and join us in the fun next time!</i></li> <li>3. <i>I'm feeling a bit stressed with the Holidays coming up. I'm looking to chat with a new social group to help.</i></li> </ol>
3	Virtual community	<ol style="list-style-type: none"> <li>1. <i>Hey Phoenix moms! I've found that sharing stories of recovery and resilience, can break down stigma and improve community.</i></li> <li>2. <i>Happy to find a new online community in this new journey! Shine On!</i></li> <li>3. <i>Good morning!! finding this new community has given me hope that staying connected and making friends will help me stay on course.</i></li> </ol>
4	In-person events	<ol style="list-style-type: none"> <li>1. <i>Tonight I attended a first of its kind, local, in-person event that brought together feminine energies in a very chill, vulnerable, and empowering way.</i></li> <li>2. <i>Thank you to everyone who showed up last night!!! It was a blast to see some returning faces, as well as a ton of new faces!</i></li> <li>3. <i>Hi! Would love to meet the Phoenix community here in my area</i></li> </ol>
5	Mindfulness (recovery)	<ol style="list-style-type: none"> <li>1. <i>Today I was mindful of my body and breath during my meditation. I also focused on being more mindful of my surroundings.</i></li> <li>2. <i>Hey! How do you bring more mindfulness into your life?</i></li> <li>3. <i>Mindfulness is such a big help in my recovery. When I start getting negative thoughts I'm able to bring my mind back to the present.</i></li> </ol>
6	Hiking to relax	<ol style="list-style-type: none"> <li>1. <i>Hey guys. Anyone use hiking to relax and get away? Let's connect!</i></li> <li>2. <i>looking for some relaxed hiking partners in [REDACTED]. Check out our hiking schedule and come relax with me!!!</i></li> </ol>
7	In-person community	<ol style="list-style-type: none"> <li>1. <i>Last night, we had a party for the wonderful folks who offer their time, energy, and passion to our community. Drop a comment and show them some love!</i></li> <li>2. <i>Hey everyone! Starting May 2nd we will be doing a month long community challenge all about meditation and mental health.</i></li> <li>3. <i>New to the sober community! Need to find a sponsor and supportive group.</i></li> </ol>
8	Planning a hike	<ol style="list-style-type: none"> <li>1. <i>Is there a [REDACTED] hike coming up soon? I'm thinking about going in the morning but wasn't sure if there was already a planned hike. Thanks!</i></li> <li>2. <i>Come join us for a hike this Saturday.</i></li> <li>3. <i>We hike regularly on Sundays if anyone is interested let me know!!!!</i></li> </ol>
9	Relapse and recovery	<ol style="list-style-type: none"> <li>1. <i>Trying to accomplish sobriety momentum. Having trouble getting started after a recent relapse.</i></li> <li>2. <i>Unfortunately I relapsed. Today is my first serious step in my road to recovery.</i></li> <li>3. <i>I lost my sobriety on Thursday &amp; I relapsed again yesterday.</i></li> </ol>
10	Mindfulness (yoga)	<ol style="list-style-type: none"> <li>1. <i>I am interested in learning to teach yoga and meditate any recommendations on how to do this? Thx.</i></li> <li>2. <i>Yoga and meditation was awesome. Motivational and totally worthwhile. Appreciate all the volunteers, staff and instructors. Thank you.</i></li> <li>3. <i>However, I am back to it today, I did about 20 minutes, very gentle. My anxiety was really high today, and yoga and meditation really did help.</i></li> </ol>

<sup>a</sup>Note that the discourse appears to contain a mix of app-member and app-volunteer dialog. All excerpts reported in this table have been slightly abridged to uphold the privacy and anonymity of app members by removing personal details, locations, and graphic depictions of substance use.

The excerpts in Table 2 highlight representative themes, discussions, or comments within our sample. These excerpts were identified using a sorting function that categorized app posts into one of the 10 topics identified by the BER-Topic tool. These quotes demonstrate a distinct and recurring theme of community support. This support manifests through encouragement, the promotion of online activities, and in-person events, including hiking, physical activity, mindfulness, meditation, and other similar activities that promote recovery through movement and community engagement.

## RQ2. What Do Compound Sentiment Analysis Scores Imply Regarding

## Personal and Collective Experiences of Phoenix App Members?

To further substantiate the topic modeling findings, which were used to identify recurring themes in the social media entries collected from the dataset, we applied VADER sentiment analysis to all collected entries, excluding duplicates and retaining “zero” values. Table 3 provides an overview of our findings, highlighting the overall mean VADER score for the entire sample, with a possible range of  $-0.99$  to  $0.99$ . It also includes a summary of highly positive entries—those scoring above  $0.50$ —and highly negative entries—those scoring below  $-0.50$ .

**Table 3.** Descriptive VADER<sup>a</sup> findings overall and stratified by highly negative and highly positive entries<sup>b</sup>.

Category	Mean VADER score (SD)	Median VADER score (IQR)	Quartiles of the positive VADER scores			Sample size, n (%)
			0.25	0.50	0.75	
Overall data	0.44 (0.42)	0.54 (0.83)	0.00	0.54	0.83	17,922 (100)
Highly negative posts	Below $-0.50$	— <sup>c</sup>	—	—	—	429 (2.4)
Highly positive posts	Above $0.50$	—	—	—	—	9363 (52.2)

<sup>a</sup>VADER: Valence Aware Dictionary and Sentiment Reasoner.

<sup>b</sup>Within the overall data, there is a right skew, indicating a greater presence of highly valenced, or positive, entries. The median and quartile ranges likewise suggest a high presence of “0.00” VADER values, which comprise neutrally valenced, or nonemotive, entries (eg, “Meet at the park today?”), or very short entries that were not scorable with the VADER lexicon (eg, “Hi”).

<sup>c</sup>not applicable.

The mean VADER score of our entire sample, after eliminating duplicate entries, was  $0.44$  with an SD of  $0.42$ . This suggests that, overall, valence, or affect, was highly positive with notable variation across all entries. The median and quartile regions, however, strongly imply a right-skewed distribution, implicating higher levels of positive content. When we further subdivided our sample to extract highly positive and highly negative entries, we observed that over half (52.2%) of our sample contained VADER scores greater than  $0.50$ . Conversely, only 429 entries, or 2.4% of our data, scored below  $-0.50$ .

To validate our VADER findings, we ran a Text2Emotion analysis with the overall data. Unlike VADER, which provides a single composite score that ranges from  $-0.99$  to  $0.99$ , Text2Emotion measures 5 emotion categories: (1) Happy, (2) Sad, (3) Surprise, (4) Angry, and (5) Fear. It then scores them as normalized values that sum to 1 for each entry. These scores reflect the intensity of each emotion expressed. Across all entries, “Happy” was the most prominent emotion (mean  $0.25$ ), followed by “Fear” (mean  $0.19$ ), “Sad” (mean  $0.17$ ), and “Surprise” (mean  $0.16$ ), while “Angry” was the least prominent emotion (mean  $0.04$ ). As with VADER, we noticed some potential misclassification

(eg, “That workout killed me today” scored high for “fear”). However, the findings from VADER and Text2Emotion both implicate the presence of highly positive content.

Table 4 provides 5 abridged excerpts of highly negative and highly positive posts. From these excerpts, we observed clear differences in the content. For highly positive excerpts, we observed people expressing gratitude for their sobriety and mental health; appreciation for The Phoenix, the community, and other support structures; and promotion of in-person and online events available in locations across the United States. As seen in Excerpt: “I am just so thankful to be alive today” and “Just had a great workout, make sure you stop by if you are in the [redacted] area!,” the presence of positive language with words including “thankful,” “alive,” “great,” and phrases such as “just had a great workout” contribute to higher than average VADER scores. Conversely, negative entries seemed to exhibit 1 recurring theme—individual narratives about struggles with sobriety, relapse, mental health, or voicing needs for additional support.

It is also worth noting that some negative entries concluded with words of encouragement, as seen in the excerpt: “If I can get through it, so can you.”

**Table 4.** Excerpts from highly negative and highly positive posts from the dataset<sup>a</sup>.

VADER <sup>b</sup> subset and key quotations	Score
Highly positive excerpts	
I'm grateful for new experiences in my life. I'm grateful for therapy. I'm grateful for support.	0.99
What's on my mind? I'll begin with giving thanks to the lord above for the breath in my lungs and to rebuild something better for my future.	0.99
I got to see a recovery milestone today and that means a lot to me.	0.99
I didn't go out last night, woke up this morning feeling ready to tackle the day. I wish you all a wonderful, peaceful, exciting Sunday.	0.98
[REDACTED] has teamed up with 1 Million Strong, an impact initiative that seeks to transform the way people think about addiction and recovery and The Phoenix, a National Sober Active Community, which fosters a community for those in recovery to host a sober-supportive wellness tent experience for festival-goers to cool off, relax, and enjoy alcohol-free cocktails in a comfortable environment.	0.98
Highly negative excerpts	
I know I don't want to but I need a clean and sober place to live out of the city. I go to meetings, I try and stick with the winners and yet I still find myself giving into my triggers to use. I really hate asking for help, and I find it burden others who are in recovery as well.	-0.99
Not doing well at right now. Current circumstances make it near impossible for me to get my treatments.	-0.99
I do not feel like I am up for this this. Using has taken control of what life I have left and I wish every night that the sun didn't rise. im lost and I am sad.	-0.98
Trying to get sober again after a difficult relapse that has kept me in addiction the past eight I am feeling alone and scared.	-0.96
I had to cancel attending a social gathering tonight. I felt so bad about but I am just exercising my boundaries because I do not like being in situations where people drink. Its ok to say NO when and if it feels right. A sober mind & body can hear itself.	-0.94

<sup>a</sup>Note that to uphold the privacy and confidentiality of the app members, these excerpts have been slightly abridged to remove any personally identifiable information; specific details; or graphic depictions of drug use, struggles, or relapse.

<sup>b</sup>VADER: Valence Aware Dictionary and Sentiment Reasoner.

### RQ3: How Can Data Mining Strategies Inform Future Adaptation of The Phoenix App and Member Experiences?

Our study used topic models and sentiment analysis to examine recurring themes, uses, and valences of randomly selected entries from the dataset. Using NLP, it is possible to increase the granularity of analysis to triangulate phrasings counterintuitive to recovery. Using sentiment categories and strategic queries, we attempted to allocate content that may imply highly negative experiences, warranting response or intervention. For example, we queried the data to isolate all entries containing the phrase "I need help," which also contained a negative VADER sentiment value (ie, values between -0.01 and -0.99). We also searched a few additional

phrases that, in the psychology literature, indicate personal rumination, worry, or distorted thinking, including "I am a"—a type of distorted thinking characterized by labeling oneself and "I will never"—a type of distorted thinking characterized by overgeneralizing or fortune-telling. We hypothesized that people using these phrases in a negative context reflected some distress or worry, should they appear in the dataset.

Although these queries returned very few entries, as seen in Table 5, they illustrate how some individuals used the app to seek support from other app members. In this context, seeking support reflected a general struggle with one's current recovery journey. At the same time, we observed individuals posting poetry, which may not directly reflect personal experiences with substance use, mental health, and recovery.

**Table 5.** We used strategic queries to identify posts that may indicate when someone is reaching out for help and assistance from the dataset<sup>a</sup>.

Query	N	Excerpt	VADER <sup>b</sup> score
I need help...	2	<ul style="list-style-type: none"> <li>I'm so stressed out Things could be great for a while I seem to have a handle on things but all of a sudden it's like my brain shifts quickly into the old way of thinking I don't know where it comes from in that moment of time I freeze without a thought I have lost myself I can't breathe I can't think clearly and I panic. My mind floods with thoughts of negativity I need help I want to live a meaningful life and start healing completely to move forward I'm here because I want to be I don't know what I'm doing.</li> </ul>	-0.47
I am a...	17	<ul style="list-style-type: none"> <li>I want to shout but I'm fine comes out Trapped in my own head I'm confined. I need help!</li> <li>This is getting to me I work nights and really going through depression. I got put on meds to not drink that's good-- well I am feeling very alone and can't really get to the outings. I hate feeling like my world is over and not caring about anything anymore it's not me I am an awesome person and need to get out of this rut or find some hobbies.</li> </ul>	-0.97 -0.86



Query	N	Excerpt	VADER <sup>b</sup> score
I will never...	5	• Hi I am an alcoholic and I have been drinking a lot. I do it to quiet my brain at times.	-0.54
		• Today the prettiest girl in school died	-0.95
		• Killed when my car hit another and burned to the ground	
		• I knew I was drunk yet I still had to drive	
		• Tell them I wish I could have stayed around.	
		• I know I cannot go back and give up the keys.	
		• I am not going to be there when my baby learns to walk.	
		• When she falls down and needs me to bandage her knees,	
		• I won't ever kiss away life's pain OH GOD I will never hear her talk.	
		• Yesterday marked 6 months for me. I remember not being able to make it past 4 days! Anyone have any advice to help cravings? I will NEVER go back but the cravings get so bad I get depressed.	-0.52
		Thank you!	

<sup>a</sup>While our search did not retain much data, we observed that these queries consistently returned representative data of sobriety struggles. Please note that to uphold the privacy and anonymity of Phoenix app members, these excerpts have been abridged slightly to remove personally identifiable information, locations, and graphic depictions of substance use.

<sup>b</sup>VADER: Valence Aware Dictionary and Sentiment Reasoner.

Discussion

Principal Findings

This study tested a computational NLP pipeline to analyze randomly allocated posts collected from The Phoenix app. Broadly, our findings suggest that app members use the app to promote social connection and community. These findings align with a growing body of literature indicating that mobile apps are effective systems for delivering behavior change interventions. We provide additional context below.

High Levels of Positivity and Inclusive Dialog on The Phoenix App May Demonstrate How Closed Social Media Platforms Differ From Conventional, Open Social Media Platforms Regarding Recovery

Numerous studies use VADER as a proxy for public mood or subjective well-being and as a primary measure to understand social attitudes expressed on social media during major public health events [38,40-42]. These studies typically involve the composite analysis of very large collections of posts from various social media platforms such as X, Reddit, and YouTube, often around a singular fixed topic ranging from substance use and recovery [38], reproductive rights and access [43], collective mental health [44], and others.

Although our study employs a similar use of VADER (ie, as a tool to measure the mood of text data), we observed key differences in our findings that diverge from previous studies. Indeed, prior studies using VADER sentiment analysis in both health and social media contexts sometimes report a mean VADER score near 0 [43]. Distributions of these VADER scores are commonly bimodal [45], indicating a near-equal presence of highly negative and highly positive content. When calculating a mean VADER value, this bimodal distribution can sometimes yield a mean VADER

value near 0 with a high standard deviation, making it challenging to draw reasonable conclusions about collective mood, well-being, or valence regardless of the topic—though more research on this topic is needed that addresses VADER limitations is needed.

However, in our analysis of over 17,000 recovery-related entries, we observed a mean VADER score of 0.44, reflecting a generally high presence of positive content. A more detailed review of the descriptive statistics revealed that the VADER score distribution was not bimodal but positively skewed. This finding is further supported by the relatively low number of highly negative posts (ie, those scoring lower than -0.50), which comprised just over 2% (n=429) of the corpus. In contrast, highly positive posts (ie, those scoring higher than 0.50) comprised more than 50% (n=9363) of our sample.

The general positive tone uncovered by VADER is further supported by the Text2Emotion analysis—where “happy” was the most dominant emotion—and themes uncovered in our BERTopic analysis. Some themes broadly referenced various activities and interests, while others appeared to originate from volunteers marketing specific online and in-person events to the membership. These themes may implicate online social connection, which can drive people to continue using the app as part of their recovery process.

Critically, we observed 1 topic that referenced relapse and sobriety struggles. This topic illustrates how, even if in very small quantities, people are willing to discuss their struggles openly, which we interpreted as an effort to seek support. Collectively, these topics offer evidence that, as a recovery smartphone app, curated app experiences may offer more directed discourse needed to support and sustain recovery. Future research should compare these findings with data from other similar recovery apps to provide further evidence that digital tools offer cost-effective approaches to promoting recovery regardless of geographic access.

## ***Topical Recovery Content With High Sentiment Values Illustrates the Need for Tailored and Safeguarded Social Media Experiences***

In recent years, a growing body of literature has explored social media's role in promoting social connection, particularly in relation to health and well-being [32,46-48]. The findings from that work highlight that the degree and quality of social connection tend to vary and often depend on the specific platform, medium of the platform, and context.

There are numerous examples of how social media can facilitate social connections. Studies on Facebook groups have found that the degree of connection offered was particularly beneficial for older adults [47] and non-native English speakers [49]. These groups are also the most dominant Facebook user base [49]. On X, before the 2022 acquisition by Elon Musk, people could search for communities by grouping related content through hashtags, which created pockets of conversation around specific topics. Reddit remains widely regarded as one of the strongest conventional social media platforms for fostering social connections through curated subreddits with trained moderators [50]. However, recurring issues across these platforms remain, such as low-quality health information, trolling, and exposure to harmful content counterintuitive to recovery [51].

Prior studies have explored the strengths and weaknesses of social media for recovery using conventional social media data, such as X or Twitter, including how communities form around topics such as alcohol abstinence during Dry January [52] and how others communicate about personal and familial drug use online [53]. While platforms such as Facebook, X, and Reddit have been leveraged to connect people with shared experiences, their general nature may not always yield positive experiences. This is potentially due to the overwhelming number of unique posts across a wide range of topics, a majority of which do not pertain to health and well-being, or inconsistent moderation that filters negative content.

To offset negative social media interactions, some argue that the approaches to safeguard spaces via social media-like mobile apps can create a curated social media experience tailored to specific needs, thereby enhancing both the quality of perceived social connection and perceived benefits of support communities [54,55]. The Phoenix app exhibits the characteristics of effective safeguarding through advanced monitoring and interactions between members and volunteers. The more curated nature of the app—designed with the express purpose of connecting people in recovery—allows for a more directed experience that reaches its intended audience.

Further evident in our data, which included topics related to social interaction and recovery, and sentiment analyses indicating high levels of app positivity, the app displays early signs of a supportive and balanced social media ecosystem where safeguarding is effective. By safeguarding the app, app members are protected from incongruent content that may be encountered on conventional social media platforms, creating

a safer experience that builds trust among app members and volunteers.

## ***Continued Push Toward Novel Strategies to Improve the Experiences of Online Recovery Communities and Their Impact***

Our findings suggest that digital recovery apps can promote a sense of community and belonging among people in recovery from substance use. To substantiate these observations, we applied an unsupervised NLP framework, which has been used in previous studies to contextualize how people express themselves about certain health topics or leverage online communities for social support or information seeking [32]. Collectively, these findings contribute to prior work calling for directed app experiences for behavior change, though additional research could add further context to our findings to address potential challenges associated with social apps designed with specific populations and objectives in mind. Still, additional work with and beyond the Phoenix is necessary.

Future research should strongly consider the following. First, The Phoenix app is distinct from other social media platforms in its emphasis on safeguarding and advanced monitoring to filter harmful content. While our sample did not identify any content deemed counterintuitive to recovery, further analysis of larger Phoenix datasets could provide deeper insights into the effectiveness of these processes—particularly for app members early in recovery or other vulnerable groups who are likely at a greater risk of negative effects from engaging with triggering or harmful content (ie, veterans and LGBTQIA-identifying members). From a cross-disciplinary perspective, additional work with other computer science fields, including machine learning and artificial intelligence, can help develop additional tools to directly intervene when someone posting on The Phoenix app writes in distressing tones. Qualitative interviews and codesign sessions with app members can also help explore personal experiences with the app's safety and support features, especially when compared with other mobile apps with a similar focus. Critically, any insights generated from that work would accomplish 2 tasks: (1) establish software-specific best practices for recovery-related mobile apps and (2) identify nuances in recovery communication across online versus conventional, residential programs.

Second, while our analysis points to the success of the app in fostering a supportive recovery environment, our findings have only been compared to prior studies using conventional social media data on substance recovery. Future studies should consider conducting comparative analyses of recovery forums across curated recovery apps, specifically using both qualitative and computational methods. An across-app comparison could add additional validity to our findings, particularly regarding the relative strengths of resources and tools offered by the Phoenix relative to tools and resources offered on apps from other recovery groups and organizations. These additional research directions will strongly contribute to our understanding of smartphone apps' roles and functions as promoters of positive behavior change

and lend further evidence to optimal smartphone app designs to promote sustained engagement in recovery.

## Limitations

This study is subject to limitations we hope to address in future work. First, our study involved a cross-sectional analysis of randomly allocated posts derived from The Phoenix app. While our findings offer preliminary insights into the use and purpose of the app, it is likely that our data are not entirely reflective of in-the-moment discourse that may demonstrate other app utilities and uses. For example, the dataset contained specific channels or communities directed at particular topics and experiences, including pets, art, and others. Although it is possible our data may be overrepresented by one or a few channels, it is not currently possible to assess this with the specific sample. We strongly encourage future research to use our pipeline with additional Phoenix data that removes, or at least compares, volunteers to app members. Second, we acknowledge that our NLP approach is entirely unsupervised, meaning no training data were leveraged. While this may limit the novelty of our approach, we contend that the analyses used provide a strong scaffold that is necessary to design, program, and ideally, implement more sophisticated tools for recovery, including knowledge graphs, retrieval-augmented generation systems, and fine-tuned large language models. Third, we

also acknowledge that, at the time of writing, The Phoenix app has been rebranded as NewForm. Although The Phoenix app and NewForm are similar in functionality, it is likely that newer features may result in marginal variations in our findings.

## Conclusions

This study tested a computational pipeline aimed at contextualizing the digital environment of a recovery-focused app developed by The Phoenix, a national nonprofit with over 65,000 monthly active members. Notably, our pipeline aimed to extract language, emotions, and behaviors expressed by app members to ascertain app use and its impact in fostering recovery support in a digital environment. These findings—which illustrate the broadly positive uses of an app as evidenced by the themes and emotions expressed by app members—add further evidence to the growing body of research calling for online recovery resources to address the chronic shortage of costly residential treatment options. Further, these findings likewise implicate key recovery dynamics in digital spaces—including how people in active recovery discuss their journeys, support one another, or share resources. Extracting and modeling these insights can further aid in the design of more robust tools aimed at keeping people in recovery by meeting them where they are in their journeys.

## Data Availability

The datasets generated or analyzed during this study are not publicly available due to The Phoenix's privacy policy but are available from the corresponding author on reasonable request.

## Authors' Contributions

Conceptualization: AS, BC, DV, KMH

Data curation: DV, ZS

Formal analysis: DV

Investigation: BC, DV, KMH, ZS

Methodology: DV, ZS

Project administration: BC

Resources: BC, DV

Software: DV, ZS

Supervision: BC

Validation: DV

Visualization: DV

Writing – original draft: AS, BC, DV, KMH

Writing – review and editing: AS, BC, DV, KMH, ZS

## Conflicts of Interest

DV is compensated as a research contractor by The Phoenix for their expertise in natural language processing. AS is also compensated as a research contractor by The Phoenix for their expertise in scientific writing. This contractual engagement is specific to advancing research and data analysis and does not influence the objectivity or outcomes of the research presented in this manuscript. KMH, BC, and ZS are all employees of The Phoenix.

## References

1. DiClemente CC, Norwood AEQ, Gregory WH, Travaglini L, Graydon MM, Corno CM. Consumer-centered, collaborative, and comprehensive care: the core essentials of recovery-oriented system of care. *J Addict Nurs*. 2016;27(2):94–100. [doi: [10.1097/JAN.0000000000000120](https://doi.org/10.1097/JAN.0000000000000120)] [Medline: [27272993](https://pubmed.ncbi.nlm.nih.gov/27272993/)]
2. Recovery and recovery support. Substance Abuse and Mental Health Services Administration (SAMHSA). 2024. URL: <https://www.samhsa.gov/find-help/recovery> [Accessed 2025-06-16]
3. Manderscheid RW, Ryff CD, Freeman EJ, McKnight-Eily LR, Dhingra S, Strine TW. Evolving definitions of mental illness and wellness. *Prev Chronic Dis*. Jan 2010;7(1):A19. [Medline: [20040234](https://pubmed.ncbi.nlm.nih.gov/20040234/)]

4. Ashford RD, Brown AM, Ryding R, Curtis B. Building recovery ready communities: the recovery ready ecosystem model and community framework. *Addict Res Theory*. Jan 2, 2020;28(1):1-11. [doi: [10.1080/16066359.2019.1571191](https://doi.org/10.1080/16066359.2019.1571191)]
5. Kelly JF, Stout RL, Jason LA, Fallah-Sohy N, Hoffman LA, Hoeppner BB. One-stop shopping for recovery: an investigation of participant characteristics and benefits derived from U.S. recovery community centers. *Alcohol Clin Exp Res*. Mar 2020;44(3):711-721. [doi: [10.1111/acer.14281](https://doi.org/10.1111/acer.14281)] [Medline: [32012306](https://pubmed.ncbi.nlm.nih.gov/32012306/)]
6. Key substance use and mental health indicators in the United States: results from the 2022 national survey on drug use and health. Substance Abuse and Mental Health Services Administration. 2023. URL: <https://www.samhsa.gov/data/sites/default/files/reports/rpt42731/2022-nsduh-nnr.pdf>
7. Best DW, Lubman DI. The recovery paradigm—a model of hope and change for alcohol and drug addiction. *Aust Fam Physician*. Aug 2012;41(8):593-597. [Medline: [23145400](https://pubmed.ncbi.nlm.nih.gov/23145400/)]
8. Castelpietra G, Knudsen AKS, Agardh EE, et al. The burden of mental disorders, substance use disorders and self-harm among young people in Europe, 1990-2019: findings from the Global Burden of Disease Study 2019. *Lancet Reg Health Eur*. May 2022;16:100341. [doi: [10.1016/j.lanepe.2022.100341](https://doi.org/10.1016/j.lanepe.2022.100341)] [Medline: [35392452](https://pubmed.ncbi.nlm.nih.gov/35392452/)]
9. Sarkar S, Thakur A, Sood E, Mandal P. Barriers and facilitators of addiction treatment: a qualitative study. *Int J Ment Health Addiction*. Apr 2022;20(2):672-690. [doi: [10.1007/s11469-020-00394-x](https://doi.org/10.1007/s11469-020-00394-x)]
10. Knickman J, Krishnan KRR, Pincus HA. Improving access to effective care for people who have mental health and substance use disorders. In: Finkelman EM, McClellan MB, McGinnis JM, editors. *Vital Directions for Health & Health Care: An Initiative of the National Academy of Medicine*. National Academy of Medicine; 2023:135-152. URL: <https://www.ncbi.nlm.nih.gov/books/NBK595164/> [Accessed 2025-11-26]
11. Knickman J, Krishnan R, Pincus H. Improving access to effective care for people with mental health and substance use disorders. *JAMA*. Oct 25, 2016;316(16):1647-1648. [doi: [10.1001/jama.2016.13639](https://doi.org/10.1001/jama.2016.13639)] [Medline: [27668948](https://pubmed.ncbi.nlm.nih.gov/27668948/)]
12. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. *JMIR Mhealth Uhealth*. Mar 18, 2020;8(3):e17046. [doi: [10.2196/17046](https://doi.org/10.2196/17046)] [Medline: [32186518](https://pubmed.ncbi.nlm.nih.gov/32186518/)]
13. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. *J Med Internet Res*. May 29, 2020;22(5):e17572. [doi: [10.2196/17572](https://doi.org/10.2196/17572)] [Medline: [32348255](https://pubmed.ncbi.nlm.nih.gov/32348255/)]
14. Alqahtani F, Al Khalifah G, Oyeboode O, Orji R. Apps for mental health: an evaluation of behavior change strategies and recommendations for future development. *Front Artif Intell*. 2019;2:30. [doi: [10.3389/frai.2019.00030](https://doi.org/10.3389/frai.2019.00030)] [Medline: [33733119](https://pubmed.ncbi.nlm.nih.gov/33733119/)]
15. Kang KI, Kang CM. Roles and effects of peer recovery coach intervention in the field of substance abuse: an integrative literature review. *Asian Nurs Res (Korean Soc Nurs Sci)*. Dec 2022;16(5):256-264. [doi: [10.1016/j.anr.2022.10.001](https://doi.org/10.1016/j.anr.2022.10.001)] [Medline: [36243312](https://pubmed.ncbi.nlm.nih.gov/36243312/)]
16. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. *JMIR Mhealth Uhealth*. Aug 29, 2019;7(8):e12983. [doi: [10.2196/12983](https://doi.org/10.2196/12983)] [Medline: [31469081](https://pubmed.ncbi.nlm.nih.gov/31469081/)]
17. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. *JMIR Mhealth Uhealth*. Jul 4, 2019;7(7):e11926. [doi: [10.2196/11926](https://doi.org/10.2196/11926)] [Medline: [31274112](https://pubmed.ncbi.nlm.nih.gov/31274112/)]
18. Gallagher L. Welcome to anxietyTok: an empirical review of peer support for individuals living with mental illness on social networking site TikTok. *Veritas*. 2021;3(1):24-32. [doi: [10.61372/vvrj.v3i1.2641](https://doi.org/10.61372/vvrj.v3i1.2641)]
19. Balsamo D, Bajardi P, De Francisci Morales G, Monti C, Schifanella R. The pursuit of peer support for opioid use recovery on reddit. *Proc Int AAAI Conf Web Soc Media*. 2023;17(1):12-23. [doi: [10.1609/icwsm.v17i1.22122](https://doi.org/10.1609/icwsm.v17i1.22122)]
20. Sowles SJ, Krauss MJ, Gebremedhn L, Cavazos-Rehg PA. “I feel like I’ve hit the bottom and have no idea what to do”: supportive social networking on Reddit for individuals with a desire to quit cannabis use. *Subst Abus*. 2017;38(4):477-482. [doi: [10.1080/08897077.2017.1354956](https://doi.org/10.1080/08897077.2017.1354956)] [Medline: [28704167](https://pubmed.ncbi.nlm.nih.gov/28704167/)]
21. Gauthier RP, Costello MJ, Wallace JR. “I will not drink with you today”: a topic-guided thematic analysis of addiction recovery on reddit. Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 29, 2022:1-17; New Orleans, LA, USA. Apr 29, 2022.[doi: [10.1145/3491102.3502076](https://doi.org/10.1145/3491102.3502076)]
22. Valdez D, Patterson MS. Computational analyses identify addiction help-seeking behaviors on the social networking website Reddit: insights into online social interactions and addiction support communities. *PLOS Digit Health*. Nov 2022;1(11):e0000143. [doi: [10.1371/journal.pdig.0000143](https://doi.org/10.1371/journal.pdig.0000143)] [Medline: [36812569](https://pubmed.ncbi.nlm.nih.gov/36812569/)]
23. Bergman BG, Kelly JF. Online digital recovery support services: an overview of the science and their potential to help individuals with substance use disorder during COVID-19 and beyond. *J Subst Abuse Treat*. Jan 2021;120:108152. [doi: [10.1016/j.jsat.2020.108152](https://doi.org/10.1016/j.jsat.2020.108152)] [Medline: [33129636](https://pubmed.ncbi.nlm.nih.gov/33129636/)]



24. Ashford RD, Bergman BG, Kelly JF, Curtis B. Systematic review: digital recovery support services used to support substance use disorder recovery. *Human Behav and Emerg Tech*. Jan 2020;2(1):18-32. URL: <https://onlinelibrary.wiley.com/toc/25781863/2/1> [doi: [10.1002/hbe2.148](https://doi.org/10.1002/hbe2.148)]
25. Ashford RD, Lynch K, Curtis B. Technology and social media use among patients enrolled in outpatient addiction treatment programs: cross-sectional survey study. *J Med Internet Res*. Mar 6, 2018;20(3):e84. [doi: [10.2196/jmir.9172](https://doi.org/10.2196/jmir.9172)] [Medline: [29510968](https://pubmed.ncbi.nlm.nih.gov/29510968/)]
26. Phelan C, Katz APM, Merrill JE, Jackson KM, Wray TB. Social media for recovery support for people with substance use disorder. a cross-sectional study of use patterns and motivations. *Drug Alcohol Depend Rep*. Jun 2025;15:100331. [doi: [10.1016/j.dadr.2025.100331](https://doi.org/10.1016/j.dadr.2025.100331)] [Medline: [40309381](https://pubmed.ncbi.nlm.nih.gov/40309381/)]
27. Patterson MS, Francis AN, Pew SH, et al. Exploring support provision for recovery from substance use disorder among members of a sober active community. *Sci Rep*. Mar 13, 2025;15(1):8740. [doi: [10.1038/s41598-025-92029-1](https://doi.org/10.1038/s41598-025-92029-1)] [Medline: [40082558](https://pubmed.ncbi.nlm.nih.gov/40082558/)]
28. Russell AM, Bergman BG, Colditz JB, Kelly JF, Milaham PJ, Massey PM. Using TikTok in recovery from substance use disorder. *Drug Alcohol Depend*. Dec 1, 2021;229(Pt A):109147. [doi: [10.1016/j.drugalcdep.2021.109147](https://doi.org/10.1016/j.drugalcdep.2021.109147)] [Medline: [34749199](https://pubmed.ncbi.nlm.nih.gov/34749199/)]
29. Weiss JB, Berner ES, Johnson KB, Giuse DA, Murphy BA, Lorenzi NM. Recommendations for the design, implementation and evaluation of social support in online communities, networks, and groups. *J Biomed Inform*. Dec 2013;46(6):970-976. [doi: [10.1016/j.jbi.2013.04.004](https://doi.org/10.1016/j.jbi.2013.04.004)] [Medline: [23583424](https://pubmed.ncbi.nlm.nih.gov/23583424/)]
30. Clark JL, Algae SB, Green MC. Social network sites and well-being: the role of social connection. *Curr Dir Psychol Sci*. Feb 2018;27(1):32-37. [doi: [10.1177/0963721417730833](https://doi.org/10.1177/0963721417730833)]
31. Pretorius C, Chambers D, Cowan B, Coyle D. Young people seeking help online for mental health: cross-sectional survey study. *JMIR Ment Health*. Aug 26, 2019;6(8):e13524. [doi: [10.2196/13524](https://doi.org/10.2196/13524)] [Medline: [31452519](https://pubmed.ncbi.nlm.nih.gov/31452519/)]
32. Martino J, Pegg J, Frates EP. The connection prescription: using the power of social interactions and the deep desire for connectedness to empower health and wellness. *Am J Lifestyle Med*. 2017;11(6):466-475. [doi: [10.1177/1559827615608788](https://doi.org/10.1177/1559827615608788)] [Medline: [30202372](https://pubmed.ncbi.nlm.nih.gov/30202372/)]
33. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. *arXiv*. Preprint posted online on 2022. [doi: [10.48550/arXiv.2203.05794](https://doi.org/10.48550/arXiv.2203.05794)]
34. Egger R, Yu J. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Front Sociol*. 2022;7:886498. [doi: [10.3389/fsoc.2022.886498](https://doi.org/10.3389/fsoc.2022.886498)] [Medline: [35602001](https://pubmed.ncbi.nlm.nih.gov/35602001/)]
35. Edinger A, Valdez D, Walsh-Buhi E, Bollen J. Deep learning for topical trend discovery in online discourse about Pre-Exposure Prophylaxis (PrEP). *AIDS Behav*. Feb 2023;27(2):443-453. [doi: [10.1007/s10461-022-03779-2](https://doi.org/10.1007/s10461-022-03779-2)] [Medline: [35916950](https://pubmed.ncbi.nlm.nih.gov/35916950/)]
36. Valdez D, Mena-Meléndez L, Crawford BL, Arvind A, Jozkowski KN. Online social media reactions to the overturn of Roe v. Wade: public health implications and policy insights. *Sex Res Soc Policy*. Jun 2024;21(2):616-631. [doi: [10.1007/s13178-023-00892-2](https://doi.org/10.1007/s13178-023-00892-2)]
37. Valdez D, Ten Thij M, Bathina K, Rutter LA, Bollen J. Social media insights into US mental health during the COVID-19 pandemic: longitudinal analysis of Twitter data. *J Med Internet Res*. Dec 14, 2020;22(12):e21418. [doi: [10.2196/21418](https://doi.org/10.2196/21418)] [Medline: [33284783](https://pubmed.ncbi.nlm.nih.gov/33284783/)]
38. Bathina KC, Ten Thij M, Valdez D, Rutter LA, Bollen J. Declining well-being during the COVID-19 pandemic reveals US social inequities. *PLoS ONE*. 2021;16(7):e0254114. [doi: [10.1371/journal.pone.0254114](https://doi.org/10.1371/journal.pone.0254114)] [Medline: [34237087](https://pubmed.ncbi.nlm.nih.gov/34237087/)]
39. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. *Proc Int AAAI Conf Weblogs Soc Media*. 2014;8(1):216-225. [doi: [10.1609/icwsm.v8i1.14550](https://doi.org/10.1609/icwsm.v8i1.14550)]
40. Garay J, Yap R, Sabellano MJ. An analysis on the insights of the anti-vaccine movement from social media posts using k-means clustering algorithm and VADER sentiment analyzer. *IOP Conf Ser: Mater Sci Eng*. 2019;482(1):012043. [doi: [10.1088/1757-899X/482/1/012043](https://doi.org/10.1088/1757-899X/482/1/012043)]
41. Li LY, Trivedi E, Helgren F, et al. Capturing mood dynamics through adolescent smartphone social communication. *J Psychopathol Clin Sci*. Nov 2023;132(8):1072-1084. [doi: [10.1037/abn0000855](https://doi.org/10.1037/abn0000855)] [Medline: [37498714](https://pubmed.ncbi.nlm.nih.gov/37498714/)]
42. Mayor E, Bietti LM. Twitter, time and emotions. *R Soc Open Sci*. May 26, 2021;8(5):201900. [doi: [10.1098/rsos.201900](https://doi.org/10.1098/rsos.201900)] [Medline: [34084541](https://pubmed.ncbi.nlm.nih.gov/34084541/)]
43. Valdez D, Mena-Meléndez L, Crawford BL, Jozkowski KN. Analyzing Reddit forums specific to abortion that yield diverse dialogues pertaining to medical information seeking and personal worldviews: data mining and natural language processing comparative study. *J Med Internet Res*. Feb 14, 2024;26(1):e47408. [doi: [10.2196/47408](https://doi.org/10.2196/47408)] [Medline: [38354044](https://pubmed.ncbi.nlm.nih.gov/38354044/)]
44. Rutter LA, Ten Thij M, Lorenzo-Luaces L, Valdez D, Bollen J. Negative affect variability differs between anxiety and depression on social media. *PLoS ONE*. 2024;19(2):e0272107. [doi: [10.1371/journal.pone.0272107](https://doi.org/10.1371/journal.pone.0272107)] [Medline: [38381769](https://pubmed.ncbi.nlm.nih.gov/38381769/)]



45. Otenen E, Seckin OC, Valdez D, Bollen J. Quantifying the digital phenotype of loneliness on twitter. Presented at: Proceedings of the Annual Meeting of the Cognitive Science Society; Jul 26-29, 2023; Sydney, Australia. 2023.URL: <https://escholarship.org/uc/item/2r3874sx> [Accessed 2024-09-03]
46. Stuart J, O'Donnell K, O'Donnell A, Scott R, Barber B. Online social connection as a buffer of health anxiety and isolation during COVID-19. *Cyberpsychol Behav Soc Netw*. Aug 2021;24(8):521-525. [doi: [10.1089/cyber.2020.0645](https://doi.org/10.1089/cyber.2020.0645)] [Medline: [33601944](https://pubmed.ncbi.nlm.nih.gov/33601944/)]
47. Sundar SS, Oeldorf-Hirsch A, Nussbaum J, Behr R. Retirees on Facebook: can online social networking enhance their health and wellness? CHI EA 2011. Presented at: CHI EA 2011 – 29th Annual CHI Conference on Human Factors in Computing Systems, Conference Proceedings and Extended Abstracts; 2011:2287-2292; New York, NY, USA. [doi: [10.1145/1979742.1979931](https://doi.org/10.1145/1979742.1979931)]
48. Zhang Y. College students' uses and perceptions of social networking sites for health and wellness information. *Inf Res Int Electron J* [Internet]. 2012;17(3). URL: <https://eric.ed.gov/?id=EJ984785> [Accessed 2024-09-16]
49. Germaine N. The Hispanic social media revolution: Spanish-language healthcare pages on Facebook introduction. *J Rhetor Prof Commun Glob*. 2021;3(1):3. URL: <https://docs.lib.purdue.edu/rpcg/vol3/iss1/3> [Accessed 2025-11-27]
50. Miller B. Investigating Reddit self-disclosure and confessions in relation to connectedness, social support, and life satisfaction. *J Soc Media Soc*. 2020;9(1):39-62. URL: <https://thejsms.org/index.php/JSMS/article/view/535> [Accessed 2020-05-29]
51. Baccarella CV, Wagner TF, Kietzmann JH, McCarthy IP. Social media? It's serious! understanding the dark side of social media. *Eur Manag J*. Aug 2018;36(4):431-438. [doi: [10.1016/j.emj.2018.07.002](https://doi.org/10.1016/j.emj.2018.07.002)]
52. Russell AM, Valdez D, Chiang SC, et al. Using natural language processing to explore “dry January” posts on Twitter: longitudinal infodemiology study. *J Med Internet Res*. Nov 18, 2022;24(11):e40160. [doi: [10.2196/40160](https://doi.org/10.2196/40160)] [Medline: [36343184](https://pubmed.ncbi.nlm.nih.gov/36343184/)]
53. Parker MA, Valdez D, Rao VK, Eddens KS, Agley J. Results and methodological implications of the digital epidemiology of prescription drug references among Twitter users: latent Dirichlet allocation (LDA) analyses. *J Med Internet Res*. Jul 28, 2023;25:e48405. [doi: [10.2196/48405](https://doi.org/10.2196/48405)] [Medline: [37505795](https://pubmed.ncbi.nlm.nih.gov/37505795/)]
54. Fernández Pedemonte D. Review: Why we expect more from technology and less from each other by Sherry Turkle. *AUCOM*. Dec 1, 2012;1(2):210-212. URL: <http://www.austral.edu.ar/ojs/index.php/australcomunicacion/issue/view/3> [doi: [10.26422/aucom.2012.0102.fer](https://doi.org/10.26422/aucom.2012.0102.fer)]
55. Berryman C, Ferguson CJ, Negy C. Social media use and mental health among young adults. *Psychiatr Q*. Jun 2018;89(2):307-314. [doi: [10.1007/s11126-017-9535-6](https://doi.org/10.1007/s11126-017-9535-6)] [Medline: [29090428](https://pubmed.ncbi.nlm.nih.gov/29090428/)]

## Abbreviations

**BERTopic:** Bidirectional Encoder Representation from Transformers topic modeling tool

**NLP:** natural language processing

**RQs:** research questions

**SUD:** substance use disorder

**VADER:** Valence Aware Dictionary and Sentiment Reasoner

*Edited by Lorraine Buis; peer-reviewed by Chanda Phelan Kane, Li Yan McCurdy; submitted 06.Nov.2024; accepted 19.Nov.2025; published 19.Dec.2025*

### Please cite as:

Valdez D, Heinrich KM, Collinson B, Streetman A, Sloan Z

Forging Online Community Among People in Recovery From Substance Use: Natural Language Processing and Deep-Learning Analysis of The Phoenix App User-Generated Data

*JMIR Mhealth Uhealth* 2025;13:e68438

URL: <https://mhealth.jmir.org/2025/1/e68438>

doi: [10.2196/68438](https://doi.org/10.2196/68438)

© Danny Valdez, Katie M Heinrich, Beth Collinson, Aspen Streetman, Zach Sloan. Originally published in JMIR mHealth and uHealth (<https://mhealth.jmir.org>), 19.Dec.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on <https://mhealth.jmir.org/>, as well as this copyright and license information must be included.