NeedToTalk: An Inverse Social Network For Mental Health Improvement and Well Being

Chouman Hussein† Tomokazu Matsui† Yuki Matsuda† Hirohiko Suwa† Keiichi Yasumoto†

1. Introduction

Mental health disorders present significant challenges in today's society, affecting a vast number of individuals. According to global statistics, over 322 million people experienced depression in 2017 [11]. Accurate diagnosis is, therefore, crucial for effective interventions and well-being improvement [2]. To explore ways of enhancing mental health diagnosis and proposing better treatments, the Person Specific Network (PSN)method and DATA-IN approach were initiated. PSN constructs real-time networks capturing an individual's psychological experiences using ESM data, while DATA-IN generates personalized treatment recommendations based on person-specific dynamic network models [12, 16].

Social media platforms offer potential benefits in terms of simplifying social interactions, promoting engagement and retention in services, and facilitating access to peer support networks, which can contribute to improving mental well-being [10]. However, existing social media and its usage has been linked to negative psychological outcomes such as loneliness, cyberbullying, anxiety, and depression [10, 18, 7, 6]. On the other hand, chat-based hot-lines using instant messaging platforms are a preferred and effective method for providing health promotion support and diverse health advice [5, 10]. Furthermore, well-managed anonymity can facilitate self-disclosure and thus improve mental well-being [9, 4, 17, 3].

Moreover, a study in the field of artificial intelligence and mental health improvement has found that AI-powered conversational agents using reinforcement learning show potential for personalized mental health support [13]. Considering these factors, we aim to introduce a novel social media platform called "NeedToTalk" - an inverse social network. This platform will facilitate the anonymous exchange of ideas, information, and content without fostering personal relationships or interactions. Our approach will leverage reinforcement learning and incremental learning to safeguard users' mental well-being. NeedToTalk will curate and deliver only beneficial content tailored to users' specific needs and preferences. Through ongoing adaptation, the platform intends to enhance user experiences and promote a healthier online environment.

This paper presents an inverse social network and highlights the significance of integrating innovative technologies to enhance healing and improve outcomes in the field of mental health within the environment of NeedToTalk.

† Nara Institute of Science and Technology

2. Related Work and Challenges

This section surveys existing studies related to our proposed SNS platform, NeedToTalk.

2.1 Mental Health and Existing Diagnosis

Mental health issues affect millions worldwide [11]. To address misdiagnosis challenges, Person-Specific Networks (PSN) offer personalized statistical models for psychotherapy exploration[5]. The **DATA-IN** approach, based on PSN, captures individual symptom patterns and dynamics to generate personalized treatment recommendations [16, 12]. However, PSN and DATA-IN have limitations, including data requirements, limited generalization, interpretability challenges, ethical concerns about data privacy, potential bias, and a focus on personalized strategies rather than comprehensive solutions. On the other hand, sharing recovery stories has become a central practice in mental health care, promoting empathy, understanding, and peer support[8]. Emotions play a significant role in shaping users' behaviors and content interactions[15]. Understanding users' predominant emotions can guide the offering of relevant and beneficial content.

2.2 Social Media and Mental Health

Social media platforms are crucial in providing social interaction and support for young adults with mental illness, offering access to peer support networks and opportunities for self-disclosure [5]. However, excessive social media usage, driven by features like infinite scrolling, has been linked to adverse psychological effects, including "Facebook depression" and correlations with depression, social comparisons, and body image concerns on platforms like Instagram [18, 7]. Additionally, platforms like Snapchat have been associated with lower levels of mental well-being and cyberbullying incidents [6, 1]. Addressing these issues is crucial for fostering a safer and healthier online environment [10, 18, 7, 6, 1].

2.3 Digital Anonymity and Self-Disclosure for Improved Well-being

A study about anonymity in social media found that:

- There are more online disclosure when anonymous with social ties,
- Intimacy regulation present in all conditions, stronger with real social ties,
- Negative emotions reduce disclosure in real-name conditions,
- No expected disclosure increase due to negative emotions in anonymity.

- Self-disclosure is rewarding and promotes human connection.
- Self-disclosure has therapeutic benefits.

These findings underscore the importance of self-disclosure for human connectedness and its positive impact on well-being. Consequently, social media platforms should consider accommodating anonymity preferences to enhance users' overall well-being [9].

However, it is essential to recognize that anonymity on specific social media platforms has also resulted in detrimental effects. The perceived security of anonymity may lead users to engage in self-disclosure, potentially revealing negative aspects of themselves, contributing to issues such as cyberbullying, offensive content, and a toxic culture [3, 4, 17].

2.4 AI-powered Conversational Agents for Mental Health Support

Sun et al. propose a novel approach to using conversational agents (CAs) for behavioral intervention and mental healthcare. They advocate for CAs with generative models to deliver more accessible, scalable, and personalized interventions, reducing the stigma associated with seeking help. Existing CAs often rely on pre-scripted conversations and rules, limiting their effectiveness and flexibility.

To overcome this limitation, the authors propose integrating generative models and reinforcement learning techniques. This enables continuous improvement of the generative model and the agent based on human feedback from clients and therapists during interactions. Techniques like few-shot generation, prompt engineering, and reinforcement learning from human feedback (RLHF) are used in the Human-in-the-Loop interaction.

This approach holds significant potential for sensitive fields such as mental healthcare and behavioral intervention. Further research should be conducted to explore the application of conversational agents with generative models backed by reinforcement learning in the field of mental health and well-being [13].

2.5 Challenges and Research Gaps

Despite existing research and advancements, several significant challenges and research gaps persist in the fields of mental health and social media, emphasizing the need for further exploration and innovation to improve interventions and support systems.

Diagnostic Accuracy and Efficiency:

Traditional mental health diagnosis methods often lack accuracy and efficiency, resulting in delays in providing timely interventions. Developing more precise and efficient diagnostic tools is crucial to enhance the accuracy of mental health diagnoses and improve the overall quality of care.

Diagnostic Accuracy and Efficiency:

Improving the precision and speed of mental health diagnostic tools is essential to provide timely interventions and enhance overall care quality.

User Privacy and Data Security in Digital Platforms:

Ensuring user privacy and data security on digital platforms is crucial for building trust and encouraging individuals to seek mental health support online.

Promoting a Positive Social Media Environment:

Addressing the adverse psychological outcomes of social media usage, such as loneliness, cyberbullying, anxiety, and depression, is vital. Creating supportive and positive online environments is critical to promoting mental well-being.

Balancing Anonymity and Responsible Usage:

Striking a balance between enabling anonymity for self-expression and preventing misuse or harmful behaviors is necessary to harness the benefits of online support systems.

Refinement and Ethical Considerations of AI-Powered Conversational Agents:

Further refinement of AI-powered conversational agents for mental health support is needed. Ensuring their reliability, empathy, and ethical usage, including privacy and confidentiality, is crucial.

Effectiveness of Existing Methods and Interventions:

Research gaps exist in optimizing current mental health interventions and methods to achieve better outcomes for users.

Addressing these challenges and research gaps will advance mental health, social media, and existing interventions, improving support systems for individuals and society.

3. Proposal

3.1 Introduction

In today's digital age, social media platforms have become ubiquitous, facilitating personal connections and interactions. However, the growing concerns regarding privacy, cyberbullying, and mental well-being have led to the emergence of new approaches to social networking. This proposal introduces the concept of an "Inverse Social Network" (ISN), a digital platform that enables the anonymous exchange of ideas, information, and content without forming personal relationships or interactions. The primary goal of ISN is to prioritize user privacy and mental well-being by focusing on content rather than personal connections.

3.2 Motivation

The motivation behind this proposal lies in addressing the limitations of traditional social media platforms and traditional health diagnosis. Common social media platforms often require users to share personal information and form direct connections, which can lead to privacy issues and cyberbullying. Additionally, the constant engagement and deep interactions on these platforms can negatively impact mental well-being.

ISN, represented by the proposed platform "NeedToTalk," aims to provide users with a safe and supportive space for anonymous sharing and discussions. By emphasizing content-based interactions and employing well-managed anonymity, NeedToTalk seeks to foster a positive environment for mental well-being.

3.3 Comparison of ISN and Common Social Media

An inverse social network refers to digital platforms that allow anonymous exchange of ideas, information, and content without forming personal relationships or interactions. They prioritize user privacy and prevent identifiable profiles or direct connections.

To highlight the key differences between ISN and common social media, we introduce the FRIENDe method:

Word : Interactivity

"Interactivity" refers to user engagement and interaction on all digital platforms.

Abbreviation: SNI (Social Network Interactivity)

"Social Network Interactivity" (SNI) refers to what define interactivity online.

Acronym: FRIENDe

FRIENDe is the method that is used to highlight the key difference between interactivity on comon social media and interactivity on inverse social network.

Table 1: Comparison of Common Social Media and Inverse Social Media

Feature	Common Social Media	Inverse Social Media
Friendship/Content Moderation	Users are identified by profiles. Moderation may be manual or automated.	Users remain anonymous, moderation based on user's profiling
Relationship/Data Privacy and Security	Bonds form via Shared Interest and data privacy varies.	Focus on content; privacy and security emphasized
Interactions	Users engage through various interactivity, private means.	Interaction prioritizes content over specific users
Engagement	Deep interactions involve conversations and repeated exchanges.	Users engage through content contribution and reactions
Networking	Relationships build through groups, events, and direct outreach.	Minimal personal connections, content topics prioritized
Digital Environments	Platform context significantly influences user interactions	Prioritizes anonymity and content over personal interactions.
Purpose	To create and maintain social connections, share life updates, and engage with others	To share and discuss content anonymously, without forming direct social connections

Table 1 compares "Common Social Media" and "Inverse Social Media" based on features such as user identification, content moderation, privacy emphasis, user interactions, engagement style, networking approach, digital environments, and platform purposes.

3.4 Anonymity in ISN

Anonymity is a crucial feature of ISN, ensuring user privacy and protection. In the context of digital platforms, anonymity means being unidentifiable, safeguarding an individual's identity and personal information during online interactions to prevent linkage to their real-life identities.

We classify anonymity into three levels:

Level 1: No anonymity - Real names and information accessible.

Level 2: Pseudonyms - Basic anonymity, personal info may be linked.

Level 3: Content-based anonymity - No identifiable info shared, focus on message content.

The proposed ISN platform, NeedToTalk, offers well-managed anonymity (Level 3), ensuring user safety and a positive online experience. It employs AI-powered anonymous chat to facilitate interactions while maintaining content-based anonymity.

3.5 Well-Managed Anonymity and Personalization

Well-managed anonymity, as utilized in NeedToTalk, goes beyond traditional anonymity by fostering a safe and respectful online environment. The platform tailors content and user interactions to individual needs and interests, providing a supportive space free from harmful or undesirable content. By prioritizing user privacy and mental well-being, NeedToTalk promotes positive personal growth and meaningful interactions.

This brings us to the difference between NeedToTalk and existing platforms shown in the Table below.

Table 2: Difference between NeedToTalk and existing platforms

Platform	Anonymity Level	Anonymity features	Health Support Focus	Limitations
YikYak [4]	3	No Profiles, Location-based posts	x	Cyberbullying, Limited content moderation
Whisper [17]	3	Share Secret, No personal info	x	Cyberbullying, Limited content moderation
Gravity [14]	3	Total Anonymity on the platform	x	Personal information is collected at registration, Cyber bullying
4chan [3]	3	Total Anonymity	x	Offensive content, lack of moderation, toxic culture
NeedToTalk	3	Well-managed Anonymity, Impossibility to link user ' s post to users. AI powered anonymous chat	✓	No predefined data because of Total Anonymity

3.6 Experimental Design

The proposed experimental design involves the following steps:

1. Initial Information Gathering:

First, we use a conversational agent to collect initial information from anonymous users through a pre-assessment questionnaire. Next, we understand their current emotional state. We then utilize a generative model to refine the questionnaire based on user responses, identifying the most prevalent emotion to shape the user's profile effectively.

2. Connecting Users with ISN:

We connect users in NeedToTalk, ensuring content and interactions align with their preferences. In essence, for every new state a user enters, there is a history

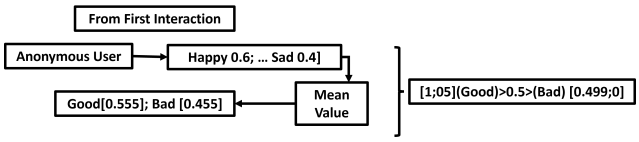

Figure 1: metrics about user state

of their previous interactions with other users or user-generated content.

This brings us to the formulation outlined below:

Let $U = \{u_0, u_1, \ldots, u_{n-1}\}$ be the set of users.

Let $S = \{s_0, s_1, \ldots, s_{n-1}\}$ be the set of states of a user, where the state of a user u is of the form $(x_{u,0}, x_{u,1}, \ldots, x_{u,k})$, with each x_j an observable interaction of u_i in NeedToTalk.

Let $C = \{c_0, c_1, \ldots, c_l\}$ be the set of consumable content in NeedToTalk.

Goal

Train an ML model that approximates u_i's profile based on s_i to perform user matching (f_1) and recommending content (f_2).

- $f_1(P_1) = \{u_i\} \subseteq U - \{u_i\}$
- $f_2(P_2) = \{c_i\} \subseteq C$

P_1-User Profile: P_1 represent the profile of a user based in their interactions and states in NeedToTalk. Each user's profile is a summary or representation of their observable interactions in the form of states, denoted as s_i

p_2-Content Profile: P_2 represents a specific content item to a user profile in NeedToTalk. The set C contains all the consumable content items availables on the paltform and each content item has its own profile, denoted as c_i

In order to effectively train our model, we need to classify both users and their content. Our approach involves categorizing all users into two main groups:

- Feeling Good Users:
  - Potential Healing Users
  - Normal Users
- Feeling Bad Users:
  - Vulnerable Users
  - Harmful Users

Similarly, we organize the content shared by users into two main categories:

- Good Content:
  - Inspirational Content
  - Entertaining Content
- Bad Content:
  - Vulnerable Content
  - Toxic Content

Analyzing user behavior and their interactions in real-time allows us to generate more precise user profiles. This classification enables us to deliver personalized and relevant content, fostering a positive and supportive environment for mental well-being.

3. Conversational Agent for Emotional Insights:

We employ a conversational agent trained with psychological data to gain insights into the user's emotions and feelings, providing better support and understanding.

4. User Classification:

We categorize users into Feeling Good (Normal and Potential Healing users) and Feeling Bad (Vulnerable and Harmful users) based on their emotions and interactions.

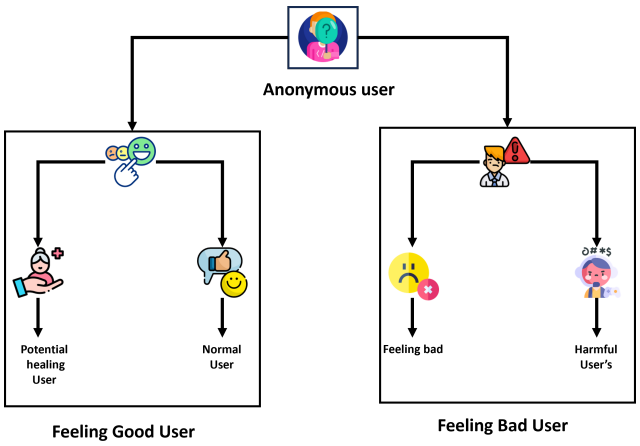

Figure 2: User's classification

5. Content Classification:

We classify shared content into Good (Normal and Inspirational) and Bad (Sad/Vulnerable and Toxic) based on their emotional impact.

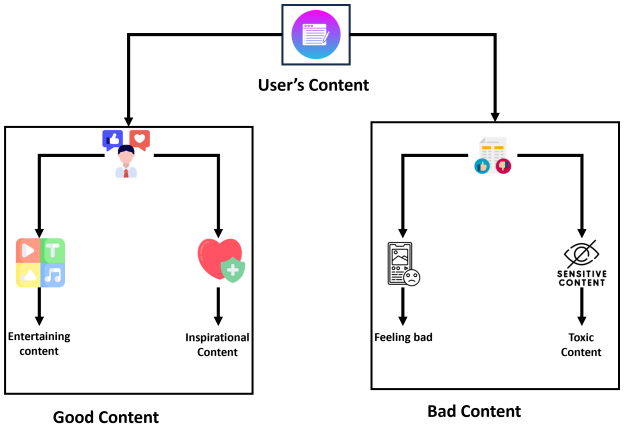

Figure 3: Content classification

We differentiate the content and the user as every kind of content and the user has specific emotions as shown in Fig. reffig:system.

Emotions Behind Our Different Users and their content

Potential Healing User	Normal User
Empathy	Open-mindedness
Compassion	Authenticity
Altruism	Optimism
Non Judgemental	Reciprocity

Harmful User	Feeling Bad
Impulsivity	Regret
Narcissism	Self-Doubt
Superiority	Need For Validation
Hatred	Negative Self-Talk

Figure 4: Emotions behind users and their content

6. User Satisfaction and Reinforcement Learning:

To assess the progress of the user's state, we propose employing reinforcement learning to define user satisfaction as a reward. By doing so, we can effectively gauge the evolution of the user's state for the better by using the following equation.

$$Q(s,a) \leftarrow (1-\alpha)\cdot Q(s,a)+\alpha\cdot(r_{t+1}+\gamma\cdot\max_{a_{t+1}} Q(s_{t+1},a_{t+1}))$$

where: $Q(s,a)$ represents the quality of action a in state s, α is the learning rate, r is the reward received after taking action a in state s, γ is the discount factor, and $\max_a Q(s,a)$ is the maximum quality of all possible actions in the next state s'.

This reinforcement learning approach allows NeedToTalk to continuously analyze users' interactions, adapt, and provide personalized support based on user satisfaction as a reward.

Our approach involves utilizing the metrics illustrated in Fig. 5 below to measure user satisfaction.

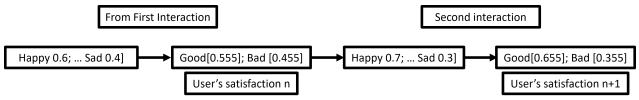

Figure 5: metrics about user's satisfaction

If the user's satisfaction$_{n+1}$ − satisfaction$_n$ is ≤ 0, no rewards are given.

3.7 Hypothesis

After getting the user state, classifying them and their content, we propose some hypothesis according to how we should match them with other users and content :

In Fig. 6, the red arrow symbolizes the priority assigned to the users, while the blue arrow represents the potential matching between the user's preferences. Based on the user's current state and the emotions behind their interaction, we propose a matching approach that caters specifically to their needs.

Fig. 7 shows how we are going to match our different users with their different content.

Personalized Content:

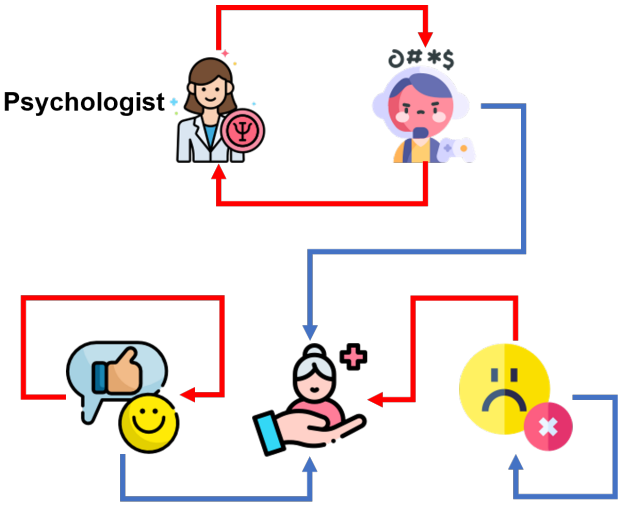

Figure 6: User's matching

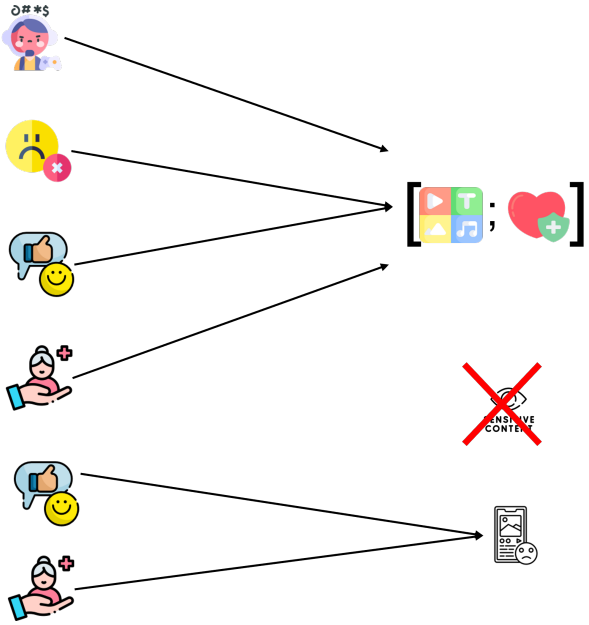

Figure 7: Content matching

We match all of the users with inspirational content and entertaining content according to the emotions behind their interaction and thus their need and interest.

Protective Measure:

We show the inoffensive content to the normal user and potential healing user in order for them to provide some insight to the user responsible for those post/content.

User Safety Priority:

To maintain the highest level of user safety, we refrain from showing inoffensive content to anyone. Our commitment is to protect our users from content that may have a negative impact on their well-being.

With those different users and content matching approach, we want to be sure to connect the users to what they need to see or may be interested in seeing by being sure that we will not match them to content or users that may hurt them.

3.8 Relation to mental health

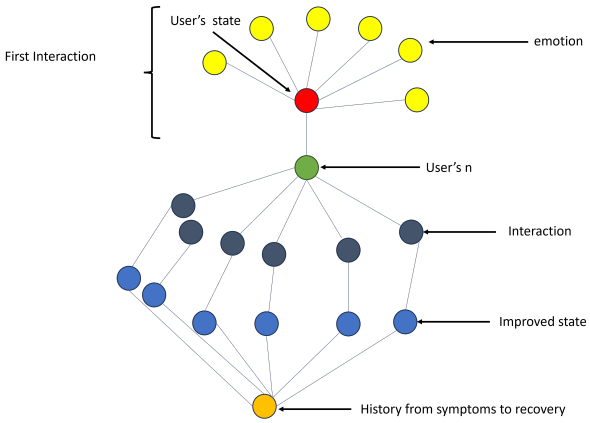

Figure 8: Graph network from first interaction to improved state

Having successfully classified our users and analyzed their content, we gained valuable insights into their interactions.

By defining a user satisfaction metric, we were able to effectively measure the improvements in their overall experience. We then utilized a graph network as you can see in the figure above to visually depict and emphasize the different stages of progress from the initial user state to their enhanced state and eventual recovery.

This approach allow us to gain a comprehensive understanding of the positive transformations our users undergo through their journey with our platform.

4. Conclusion

In this paper, we have introduced "NeedToTalk," an "Inverse Social Network" (ISN) that prioritizes user privacy and mental well-being. Unlike traditional social media platforms, NeedToTalk focuses on content-based interactions without personal relationships, fostering a safe online environment through well-managed anonymity. We proposed an experimental design that involves user and content classification, emotional insights using a CA, user satisfaction measurement with reinforcement learning, and personalized content matching. The platform's emphasis on mental health includes identifying user emotions, connecting them with relevant content, and visualizing their progress through graph networks.

Through our proposed approach, we expect to find several key outcomes. First, we anticipate that NeedToTalk's emphasis on content-based interactions and anonymity will lead to increased user privacy and reduced negative effects associated with deep personal connections on traditional social media platforms. Second, by employing reinforcement learning to measure user satisfaction and connecting users with personalized content, we hope to see enhanced user satisfaction and positive engagement within the platform. Lastly, visualizing user progress through graph networks can highlight positive transformations and eventual recovery experienced by users as they interact within NeedToTalk.

There are several avenues for future work and improvements. One area of focus would be refining user matching algorithms to ensure users are consistently connected with content and other users that are most relevant to their needs and interests. Additionally, integrating advanced emotional insights, leveraging natural language processing techniques and emotional AI, would allow for more nuanced and accurate emotional analysis of user interactions. Conducting longitudinal studies to assess the long-term impact of using NeedToTalk on user mental well-being and collaborating with mental health professionals to enhance the platform's support features are also essential steps in making a meaningful impact on the lives of our users.

References

[1] Adesoji Ademiluyi, Chuqin Li, and Albert Park. Implications and preventions of cyberbullying and social exclusion in social media: Systematic review. *JMIR Form Res*, 2022.

[2] Gaan Akers. The impact of mental health misdiagnosis. Hillside Atlanta, 2019.

[3] Michael S. Bernstein, Andres Monroy-Hernandez, Drew Harry, Paul Andre, Katrina Panovich, and Greg Vargas. 4chan and /b/: An analysis of anonymity and ephemerality in a large online community. *ICWSM*, 2011.

[4] Erik W. Black, Kelsey Mezzina, and Lindsay A. Thompson. Anonymous social media – understanding the content and context of yik yak. *Computer in Human Behavior*, 2016.

[5] Carinne Brody, Alaina Star, and Jasmine Tran. Chat-based hotlines for health promotion: a systematic review. *mHealth*, pages 6–12, 2020.

[6] Tasha R. Dunn and Michael R. Langlais. " oh, snap! " : A mixed-methods approach to analyzing the dark side of snapchat. *The journal of social media in society*, 2020.

[7] Lien Faelens, Kristof Hoorelbeke, Ruben Cambier, Jill van Put, Eowyn Van de Putte, Rudi De Raedt, and Ernst H.W. Koster Statista. The relationship between instagram use and indicators of mental health: A systematic review. *Computers in Human Behavior Reports*, pages 4–15, 2021.

[8] Llewellyn-Beardsley J, Rennick-Egglestone S, Callard F, Crawford P, Farkas M, Hui A, Manley D, McGranahan R, Pollock K, Ramsay A, Sælør KT, Wright N, and Slade M. Characteristics of mental health recovery narratives: Systematic review and narrative synthesis. *PLoS One*, 2019.

[9] Xiao Ma, Jeff Hancock, and Mor Naaman. Anonymity, intimacy and self-disclosure in social media. Association for Computing Machinery, 2016.

[10] John A. Naslund, Ameya Bondre, John Torous, and Kelly A. Aschbrenner. Social media and mental health: Benefits, risks, and opportunities for research and practice. *Journal of Technology in Behavioral Science*, pages 3–12, 2020.

[11] E. Orti, J.L. Bredas, and C. Clarisse. Depression and other common mental disorders. *Global health Estimates*, 2017.

[12] Julian A. Rubela, Aron J. Fisher, Kristin Husen, and Wolfgang Lutza. Translating person-specific network models into personalized treatments: Development and demonstration of the dynamic assessment treatment algorithm for individual networks (data-in). *Psychotherapy and Psychosomatics*, 2018.

[13] Xin Sun, Jos A. Bosch, Jan De Wit, and Emiel Krahmer. Human-in-the-loop interaction for continuously improving generative model in conversational agent for behavioral intervention. Association for Computing Machinery, 2023.

[14] Sungrove. Gravity とは?本当に「優しい sns」?新機能の使い方も解説. Sungrove “最先端の Web マーケティングを発信するメディア”, 2022.

[15] UWA. The science of emotion: Exploring the basics of emotional psychology. Psychology and Counseling News, 2019.

[16] Lino von Klipstein, Harriëtte Riese, Date C. van der Veen, Michelle N. Servaas, and Robert A. Schoevers. Using person-specific networks in psychotherapy: challenges, limitations, and how we could use them anyway. *BMC medicine*, 2020.

[17] Gang Wang, Bolun Wang, Tianyi Wang, Ana Nika, Haitao Zheng, and Ben Y. Zhao. Whispers in the dark: Analysis of an anonymous social network. Association for Computing Machinery, 2014.

[18] Sunkyung Yoon, Mary Kleinman, Jessica Mertz, and Michael Brannick. Is social network site usage related to depression? a meta-analysis of facebook–depression relations. *Journal of Affective Disorders*, pages 3–12, 2019.