University of South Carolina

Scholar Commons

2021

"Who can help me?": Knowledge Infused Matching of Support Seekers and Support Providers during COVID-19 on Reddit

Manas Gaur
University of South Carolina - Columbia, mgaur@email.sc.edu

Kaushik Roy
University of South Carolina - Columbia, kaushikr@email.sc.edu

Aditya Sharma
LNMIIT - India, 17ucs011@lnmiit.ac.in

Biplav Srivastava
University of South Carolina - Columbia, biplav.s@sc.edu

Amit Sheth
University of South Carolina - Columbia, amit@sc.edu

Follow this and additional works at: https://scholarcommons.sc.edu/aii_fac_pub

Part of the Artificial Intelligence and Robotics Commons, Databases and Information Systems Commons, Data Science Commons, Digital Humanities Commons, Graphics and Human Computer Interfaces Commons, Health Information Technology Commons, Mental and Social Health Commons, and the Social and Behavioral Sciences Commons

Publication Info

"Who can help me?": Knowledge Infused Matching of Support Seekers and Support Providers during COVID-19 on Reddit

Manas Gaur*, Kaushik Roy*, Aditya Sharma†, Biplav Srivastava*, Amit Sheth*

* AI Institute, University of South Carolina, USA {mgaur, kaushikr}@email.sc.edu {biplav.s, amit}@sc.edu

† LNMIIT, India 17ucs011@lnmiit.ac.in

Abstract—During the ongoing COVID-19 crisis, subreddits on Reddit, such as r/Coronavirus saw a rapid growth in user's requests for help (support seekers - SSs) including individuals with varying professions and experiences with diverse perspectives on care (support providers - SPs). Currently, knowledgeable human moderators match an SS with a user with relevant experience, i.e, an SP on these subreddits. This unscalable process defers timely care. We present a medical knowledge-infused approach to efficient matching of SS and SPs validated by experts for the users affected by anxiety and depression, in the context of with COVID-19. After matching, each SP to an SS labeled as either supportive, informative, or similar (sharing experiences) using the principles of natural language inference. Evaluation by 21 domain experts indicates the efficacy of incorporated knowledge and shows the efficacy the matching system.

Index Terms—COVID-19, Social Roles, Support Seekers, Support Providers, Online Mental Health Communities, Medical Knowledge Bases, Deep Semantic Clustering, Social Computing

I. INTRODUCTION

Within two months of the novel coronavirus pandemic, Reddit' subreddit *r/Coronavirus'* subscribers jumped from 2,000 to a staggering ∼2 Million with "Reddit potentially being the internet's best support group" [1] [1]. Subreddits *r/Coronavirus* and *r/covid19_support* are a significant source of information for those coping with Anxiety. Reddit mirrors people's thoughts and actions, invaluable in crafting responsive plans. Presently, these subreddits have 64 moderators with diverse backgrounds: (a) academics in genomic science, infectious diseases, virology, and tuberculosis, (b) healthcare workers including nurses, general practitioners, internal medicine specialists, and mental health specialists, and (c) public health personnel and epidemiologists (see Figure 1). The enormous set of users in these communities has overwhelmed moderators. Worse, users drift in from other mental health subreddits (*r/Depression, r/Anxiety, r/Opiates, r/SuicideWatch*) to *r/Coronavirus* seeking support for their deteriorating mental health conditions due to COVID-19 [2]. Broadly, the conversations on *r/Coronavirus* surround major events such as {business closure, school closure, lockdown, shelter-in-place, hospitalization} (also non-pharmaceutical interventions (NPIs)), and activities of daily living. We use mental health lexicons from the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [3], depression using
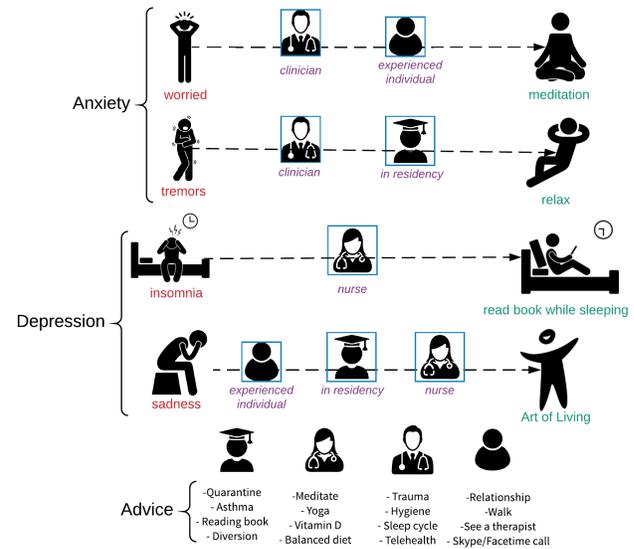

Fig. 1: Illustration of a scenario in which users suffering from Anxiety and Depression (SSs) seek help. The moderators endeavor to connect them to the users with the relevant experience (SPs) to address the SS's needs. The moderators may themselves provide advice.

Patient Health Questionnaire-9 (PHQ-9) [4], anxiety [5], and information on NPIs to measure the impact of COVID-19 on Reddit users' mental health.

Users and moderators started the *r/covid19_support* community to provide emotional support specific to COVID-19. The community received significant traction during February to March 2020, with user counts reaching 20K. One nurse moderator, referred concerned users on the *r/Coronavirus* subreddit to *r/covid19_support* for support on Anxiety and Depression that is virus-related[2]. However, the task of associating a support seeker (SS) with a potential support provider (SP) within these dynamic communities dwarfs the limited number of moderators.

Can fine-grained knowledge about users improve matching of SS and SPs on r/Coronavirus and r/covid19_support

subreddits?

We used data from *r/Coronavirus* and *r/covid19_support* communities during the first wave of the COVID-19 - when people started experiencing symptoms of mental illness. To detect topics and issues that reference Anxiety and Depression, we used PHQ-9[3], anxiety[4], and DSM-5[5] lexicons. These medical knowledge resources support contextual embeddings of SS's posts. We summarize our key contributions as follows:

1) **Dataset**: We include the characteristics of support seekers and providers, along with the situation. On the crawled 5.3 Million posts and comments from COVID-19-related subreddits, we used psycholinguistic [6] and pandemic-related knowledge on events to create a working dataset for matching SSs with SPs. We also leverage a general gold-standard Reddit dataset on supportive and support-seeking behaviors (See Section III).

2) **Method**: Automated matching of SS with SPs is challenging because their content generates dissimilar embedding. Health-specific markers of SS differ from SP. Hence, we adapt a model that emphasizes contrastive pairing and allows knowledge-infusion. We use a convolutional siamese network model to match SS and SPs.

3) **Quantitative evaluation**: Reddit conversations for an SS and SP match resemble Natural Language Inference (NLI) outcomes of entailment, contradiction and neutral. Hence, we evaluate the matches using the NLI paradigm by labeling the SPs as similar(entailment), supportive(contradiction), and informative(neutral) to an SS problem.

4) **Qualitative Evaluation**: We evaluated the best performing method qualitatively with a cohort of 21 Domain Experts (DEs).To the best of our knowledge, this is the first study that leverages diverse domain-specific knowledge to match online support seekers and providers.

In the pandemic context, this ability assists the strategic *distribution of limited resources*.

II. RELATED WORK

The stigma surrounding mental health encourages patients to seek peer-support anonymously on community-based platforms, such as Reddit or Talk life [7]. Powell et al's 12-item general population survey demonstrated that individuals with mental illness seek social support through sensitive self-disclosure [8]. Andalibi et al. categorized user behaviors into well-established classes: supportive and unsupportive [9]. O'Leary et al. suggested the importance of user expectations, roles, risk, and clinical knowledge for meeting care demands online [10] with peer support. Gillani et al., showed that a simple reframing of irrational thoughts through support provider perspectives could bring positive cognitive change to those suffering from mental stress [11]. However, prior research did not employ contextualization and abstraction

of social media content, a requirement for capturing health-specific markers of an SS user and match with an SP, who can provide relevant help [12]. We build and improve upon related research [13], [14] on SS and SP role identification by forming dynamic support groups using psycholinguistic and clinical knowledge braided with deep learning.

III. EXPLORATORY DATA ANALYSIS

We collected nearly 5.3 Million posts and comments from nearly 450K users in *r/Coronavirus* subreddit and 48K posts/comments from 6.9K users in *r/covid19_support* subreddit (see Table I).

Timeframe	*r/covid19_support*=February 22 to May 31, 2020 *r/Coronavirus* = January 13 to May 13, 2020
#Posts in subreddits	*r/Coronavirus*= 5.3M, *r/covid19_support*=48K
#Users in subreddits	*r/Coronavirus*=523K, *r/covid19_support*=7K
#Users per mental health condition	Anxiety=8K and Anxiety-SS = 6.8K, Depression= 4K and Depression-SS=2.7K
#Posts per COVID-19-related events	Business Closure=14.7K, School Closure=13K, Lockdown= 27.6K, shelter-in-place=8K, Hospitalization=1K

TABLE I: Reddit statistics for posts and users. The time difference between r/Coronavirus and r/covid19_support subreddits is due to the delayed opening of r/covid19_support for help-seeking users in the r/Coronavirus subreddit.

The DEs involved in this study noted that (1) Typical comments and posts in *r/covid19_support* subreddit are either supportive/helpful, and (2) There is a fair distribution of problem-focused and informative comments/posts in *r/Coronavirus* subreddit, but rarely a supportive post/comment. Thus, we merged the post/comments from *r/Coronavirus* and *r/covid19_support* subreddits. From this large-scale data, we constructed a subset of users with their posts/comments specific to mental health (Anxiety and Depression) and events in COVID-19 (e.g., business closure, school closure). Figure 2 shows the entire exploratory data analysis pipeline

Event-based Filtering: School closure, business closure, lockdown, shelter-in-place, and hospitalization, and general COVID-19 conversation during COVID-19 likely had a global psychological impact (see Table II). We filtered posts and comments in the raw data through soft matches with COVID-19-related events using embedding with a cosine similarity threshold over 0.8. Subsequently we filtered posts and comments for concerns regarding Anxiety and Depression through SS or SP behavior.
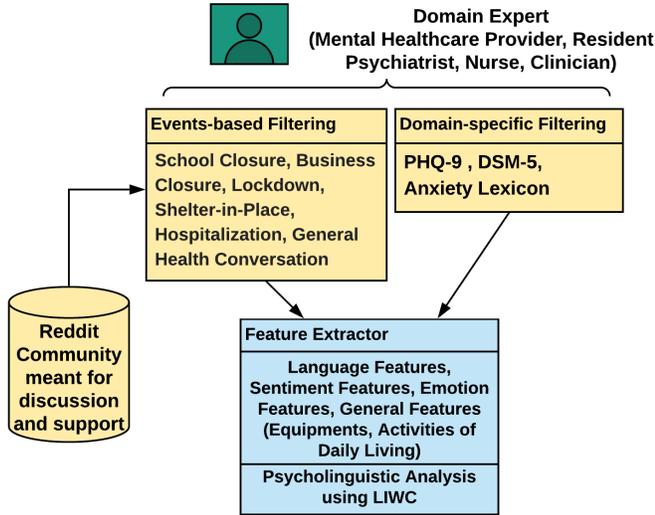
Fig. 2: Feature extraction pipeline distilling noisy Reddit posts by significant COVID-19-events, and clinical terms on anxiety and depression; subsequently we generate language, sentiment, emotion, ADLs, and psycholinguistic feature vectors along with content to representation SS and SPs

Psy/C/Ev	D	A	Psy/C/Ev	D	A
Emotional	0.39	0.5	**Social**	0.24	0.6
Biological	0.23	0.48	**Cognitive**	0.25	0.53
Future	0.001	0.23	**Modals**	0.31	0.64
Inst ADLs	0.21	0.4	*Basic ADLs*	0.18	0.3
Equipment	0.22	0.33	*Sch Closure*	0.21	0.23
Bus Closure	0.2	0.31	*Lockdown*	0.22	0.29
Shelter	0.2	0.37	*Hospital*	0.24	0.4

TABLE II: Correlations between depression (D) or anxiety (A) posts and **psycholinguistic features (Psy)**, *COVID-19 related terms (C)* (bold + italics) and *COVID-19 related events (Ev)*

Identify Users with Anxiety and/or Depression: We used an anxiety lexicon and a negation detection method to tag complex posts or comments in the dataset correctly. The lexicon captured varied expressions of anxiety, such as anxiety, anxiousness, anxious, agita, agitation, prozac, sweating, and panic attacks. For instance, the following post: "Then others that insisted that what I have is depression even though *manic episodes* aren't *characteristic of depression*. I dread having to retread all this again because the clinic where I get my mental health addressed is closing down due to loss in business caused by the pandemic" was tagged as "Business Closure with Anxiety" as italicized phrases appear in the semantic lexicon. Negation detection identified "aren't," as the negation, making this sentence as "not depression," and tagging it as "anxiety." Table II notes that hospitalization and lockdown events caused relatively higher anxiety levels in individuals compared to business closure, school closure, and shelter-in-place. For example, posts like "*Need help. Have a friend who lives alone who is now suicidal from the isolation and anxiety, and already had depression. I've asked her to come to my house for the shelter in place, but she doesn't want to*" were tagged as "Shelter-in-place with Depression and Anxiety."

Feature Extraction: We discuss two types of features that characterize SSs and SPs:

a) Psycholinguistic features: Given tagged users with the most relevant mental health condition (anxiety or depression), we extracted lexical features describing concerns surrounding cognitive, emotional (positive emotions, negative emotions, affect, anger, sad), biological (sexual, body, ingest, health), Focus Future, and social processes (Social, Family, Friend).

LIWC[6] (linguistic Inquiry and Word Count) provides a comprehensive categorized dictionary of words that people use words in their daily lives. This provides rich information about their beliefs, fears, thinking patterns, social relationships, and personalities in crisis [6].

b) COVID-19 features: Modals (epistemic and dynamic), Instrumental Activities of Daily Livings (Inst ADLs), Basic ADLs, and Equipment are additional features in conversations specific to COVID-19. Instrumental ADLs include "moving within the community," "preparing meals," "using telephone for communication," "cleaning and maintaining the house," "taking prescribed medication," and "managing money". Basic ADLs comprise "personal hygiene," "toilet hygiene," "functional mobility (e.g., getting in and out of bed), and self-feeding. LabMT, a Mechanical Turk emotion assessment tool, provided a real-valued score for emotional features [7].

General Findings: We note the following (see Table II). Users who expressed depression were affected by the disruption in social processes, epistemic modality (*described as pandemic uncertainty with words like may or might* and dynamic modality (*described as possible or necessary conditions expressed by words like can or will in a sentence*) (see Table II). COVID-19 has impacted ADLs. The effect is greater for users with depression compared to users with anxiety. Depressed individuals used words that reflected concerns on their post-COVID-19 futures, which we capture by mapping them to term under *Focus Future* category in LIWC [8]. The association of such words was higher in depressed individuals compared to users with anxiety. Questions or concerns such as adjustment with traumatic events, how to recover from anorexia, having difficulty in thinking, and poor expression of thoughts scored higher on the cognitive process category for users with depression than with anxiety. The frequent usage of phrases similar to "disruption in sleep", "low body vitals", and "sexual words" suggested an impact on the biological processes category during COVID-19. This appeared equally prominent in users with depression and anxiety. Further, conversations mentioning isolation, social distancing, and

loneliness mainly came from users with depression compared to those with anxiety. Emotional processes of both users with depression and anxiety were equally affected, showing significantly high scores in anger, frustration, and sadness. Situations like "unable to get prescribed drugs" lead to more anxiety concerns than depression concerns. We performed statistical significance testing with Bonferroni-correction (p-value<0.05), showing that features in table II) are discriminative for anxiety-depression.

SS and SP classification: To classify users as either SS or SP, we used a labeled dataset from a recent study on the assessment of suicide risk. The reported Krippendorff agreement was 0.79, sufficient to be a gold standard [9] [4], [15]. Over the annotated dataset, we experimented with various pipelines, generating contextualized embedding followed by a classifier to predict SP or not SP. For generating embeddings of the posts/comments, we used: ELMo [16], Universal Sentence Encoder (USE), and GPT2 Embeddings [17] and for classification, we used: Logistic Regression (LR), and Support Vector Machines. GPT2+LR is the best performing classifier for this dataset. We chose these representation learning models as they are State-of-the-Art and can generate context-aware embeddings. Furthermore, we do not use BERT as it requires input chunking due to its max token length of 512 [18]. We used the GPT2+LR model to generate probabilities of a user being SS in our dataset. We do not label SPs as we know the ground truth SPs annotated by DEs to be 108. We labeled approximately 9.5K (6.8K Anxiety and 2.7K Depression, Table I) users as SS and found that the SS:SP ratio of $\sim 10K$:100 reflects the true distribution on Reddit.

Methods	Precision		Recall		F1-Score	
	SS	SP	SS	SP	SS	SP
Without KI (Content)	0.79	0.48	0.79	0.48	0.79	0.48
With KI (Psy, $P_{SS,SP}$)	0.88	0.68	0.62	**0.90**	0.73	0.77
With KI (Psy, $P_{SS,SP}$, COVID-19)	0.72	0.88	**0.94**	0.55	0.82	0.68
With KI (Content, Psy, $P_{SS,SP}$, COVID-19)	**0.89**	**0.89**	0.92	0.86	**0.90**	**0.87**

TABLE III: An Ablation study with KI with three possible configurations: Psy: Psycholinguistic Knowledge, $P_{SS,SP}$ = Pr(X={SS,SP}). With Psycholinguistic knowledge, the recall on SPs increases. However, SS recall rises with the addition of COVID-19 knowledge. Combining the two improves performance overall. Ablation study improvements are bold-faced.

IV. KNOWLEDGE INFUSED MATCH PREDICTION

In the proposed approach, Knowledge-infused Match (KI-Match), we train a Convolutional Siamese Network Model with contrastive loss to predict matches, computed using the knowledge features and the GPT-2 encoding of the posts, using low dimensional representations of the users for details on the architecture please see [19]

$$CoSim(SS, SP) - CoSim(SS, \overline{SP}) + \alpha \leq 0$$

Where SS is the support seeker post, SP is a relevant Support Provider, and $\overline{SP}$ is a non-relevant support provider. *CoSim* is the cosine similarity between the two data points, and α is the margin. Specifically, we pass a concatenated vector of the GPT-2 encoded post vector and the features extracted, namely the pyscholinguistic features (Psy), and COVID-19 features into the siamese network blocks to construct low dimensional representations. Each Siamese Network block is a Convolutional Neural Network, and the supervision signal is binary, i.e. Label 1 denotes a good match and 0 a bad one.

V. RESULTS AND ANALYSIS

Quantitative Evaluation: After hyper-parameter tuning, we compared the best performing Siamese Network models using precision, average recall, and f1-score (See Table III). GPT-2 clustering without KI [20] is above the different configurations of knowledge models. KI methods with different forms of knowledge in the form of psycholinguistic features, COVID-19 features, and probability of support provider/seeker improves the accuracy of match prediction.

Qualitative Evaluation: Previous work improved the learned representations of neural network models using NLI methods. Instead, we use NLI in labeling the SPs matched to an SS. In general, NLI compares a premise statement with a hypothesis statement [21]–[24]. NLI outputs labels on the hypothesis as, *entailment* - entailed from the premise, *contradiction* - contradicting the premise and *neutral* - neutral to the premise. We consider the premise as the SS post and the hypothesis as the SP post. *Supportive* posts from SP users should contradict the SS posts, *Similar* SP users will be entailed from the SS posts, and *Informative* SP users will be neutral to the SS's post. Using this labeling scheme and a pre-trained and fine-tuned RoBERTa transformer model, we determined the NLI labels for SPs (see Table IV). Further, a cohort-based assessment of matched SS and SPs (also labeled) from 21 domain experts showed 75% relevance agreement (they consider 3 out of 4 recommended SP users as relevant to an SS user) with the KI-Match prediction model. Each expert was asked: (1) How many **Users** can provide support/help/information of use to the **Problem User**. (2) What is your confidence in the response? on a random set of 21 SSs expressing anxiety, 21 SSs expressing depression and each having 4 recommended SPs per SS. As Table V shows, Faculty, Ph.D. students and undergraduates (UG) gave a confidence score of 7 out of 10, whereas medical professionals gave a confidence score of 8/10 in predicted

Problems	Responses
I am not sleeping much anymore. **Anxiety** is pretty high for the stability of the world and the future of trust. Probably need to take up drinking or something...	*Supportive*: Giving up is in your control. Exercise can be lots of different things and a way to help anxiety.
	Informative: Anxiety is quite inducing. A good time to learn relaxation techniques
	Similar: I hear you. Myself and other friends with kids are going through similar anxiety right now [...]
Married with a supportive husband but my serious health issues including **depression and PTSD** has made me feel as if I am losing everything [...]	*Supportive*:Tough position [...] kind of relate [...] don't think marriage is dying [...] advice would be to meditate, prioritize, and act patiently. Be positive [...]
	Similar: My MDD is affecting my married life. I am an outdoor enthusiast and so is my husband. My health concerns keep pulling him down [...] wanted him to let me go.

TABLE IV: Examples of SS problems and SP responses categorized as *Supportive*, *Similar* or *Informative* by the NLI model: RoBERTa. It can be seen, how *Supportive* posts contradict the view in the SS's post ("*My point is anxiety is worse than death, Do not let your mind slip*"), *Similar* posts are entailed in the SS's post ("*I hear you, I am on the same boat*"), and *Informative* posts are neutral, specifying guidelines and coping mechanisms.

Cohort	C1	C2	C3	C4
Faculty (5)	2.96	7.16	2.24	6.84
MPs (5)	3.08	8.4	2.8	7.92
Ph.D. (7)	2.26	6.45	2.0	6.85
UG (4)	2.65	6.15	2.55	7.40

TABLE V: Mean #SPs selected by experts from 4 recommendations per SS having anxiety problems (**C1**). Their confidence (on scale 1 to 10) in selection is reported in **C2**. Likewise for Depression, mean #SPs selected by experts from 4 recommendations per SS and their confidence is reported in **C3** and **C4** respectively. MPs: Medical Professionals.

matches. Thus, some variance remains in the shallow infusion of knowledge occurring through psycholinguistic and language-oriented features.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

Limitations of this research include: (1) The comparison of the quantitative aspects of SS-SP matching to NLI definitions is analogous to common sense understanding. However, surveying members who occupy such roles on Reddit would allow us to compare with user perception. (2) Infusing clinical and psycho-social knowledge could be improved through attention-based methods. (3) Mental health-specific questionnaires and expert-QAs could be used to match SS with SPs with explanations. These require behavioral therapists' supervision.

ACKNOWLEDGEMENT

REFERENCES

[1] D. M. Low, L. Rumker, T. Talkar, J. Torous, G. Cecchi, and S. S. Ghosh, "Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study," *Journal of medical Internet research*, 2020.

[2] A. Alambo, S. Padhee, T. Banerjee, and K. Thirunarayan, "Covid-19 and mental health/substance use disorders on reddit: A longitudinal study," *arXiv preprint arXiv:2011.10518*, 2020.

[3] M. Gaur, U. Kursuncu, A. Alambo, A. Sheth, R. Daniulaityte, K. Thirunarayan, and J. Pathak, "" let me tell you about your mental health!" contextualized classification of reddit posts to dsm-5 for web-based intervention," in *Proc. of ACM CIKM*, 2018.

[4] M. Gaur, V. Aribandi, A. Alambo, U. Kursuncu, K. Thirunarayan, J. Beich, J. Pathak, and A. Sheth, "Characterization of time-variant and time-invariant assessment of suicidality on reddit using c-ssrs," *arXiv preprint arXiv:2104.04140*, 2021.

[5] L. Rheault, "Expressions of anxiety in political texts," in *Proc. of EMNLP Workshop on NLP and CSS*, 2016.

[6] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc," *Mahway: Lawrence Erlbaum Associates*, 2001.

[7] Y. Pruksachatkun, S. R. Pendse, and A. Sharma, "Moments of change: Analyzing peer-based cognitive support in online mental health forums," in *Proc. of CHI*, 2019.

[8] J. Powell and A. Clarke, "Internet information-seeking in mental health: population survey," *The British Journal of Psychiatry*, 2006.

[9] N. Andalibi, P. Ozturk, and A. Forte, "Sensitive self-disclosures, responses, and social support on instagram: the case of# depression," in *Proc. of ACM CSCW*, 2017.

[10] K. O'Leary, A. Bhattacharya, S. A. Munson, J. O. Wobbrock, and W. Pratt, "Design opportunities for mental health peer support technologies," in *Proc. ACM CSCW*, 2017.

[11] N. Gillani, A. Yuan, M. Saveski, S. Vosoughi, and D. Roy, "Me, my echo chamber, and i: introspection on social media polarization," in *Proc. World Wide Web Conference*, 2018, pp. 823–831.

[12] M. Gaur, K. Faldu, and A. Sheth, "Semantics of the black-box: Can knowledge graphs help make deep learning systems more interpretable and explainable?" *IEEE Internet Computing*, vol. 25, no. 1, pp. 51–59, 2021.

[13] H. Purohit, A. Hampton, S. Bhatt, V. L. Shalin, A. P. Sheth, and J. M. Flach, "Identifying seekers and suppliers in social media communities to support crisis coordination," *CSCW*, 2014.

[14] D. Yang, R. E. Kraut, T. Smith, E. Mayfield, and D. Jurafsky, "Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities," in *Proc. of CHI*, 2019.

[15] M. Gaur, A. Alambo, J. P. Sain, U. Kursuncu, K. Thirunarayan, R. Kavuluru, A. Sheth, R. Welton, and J. Pathak, "Knowledge-aware assessment of severity of suicide risk for early intervention," in *Proc. of The Web Conference*, 2019.

[16] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," *arXiv preprint arXiv:1802.05365*, 2018.

[17] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar *et al.*, "Universal sentence encoder for english," in *Proc. of EMNLP*, 2018.

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[19] L. A. Vilhagra, E. R. Fernandes, and B. M. Nogueira, "Textcsn: a semi-supervised approach for text clustering using pairwise constraints and convolutional siamese network," in *Proc. of ACM SAC*, 2020.

[20] V. Kieuvongngam, B. Tan, and Y. Niu, "Automatic text summarization of covid-19 medical research articles using bert and gpt-2," *arXiv preprint arXiv:2006.01997*, 2020.

[21] K. Sinha, P. Parthasarathi, J. Wang, R. Lowe, W. L. Hamilton, and J. Pineau, "Learning an unreferenced metric for online dialogue evaluation," *arXiv preprint arXiv:2005.00583*, 2020.

[22] A. Poliak, A. Haldar, R. Rudinger, J. E. Hu, E. Pavlick, A. S. White, and B. Van Durme, "Collecting diverse natural language inference problems for sentence representation evaluation," *arXiv preprint arXiv:1804.08207*, 2018.

[23] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes, "Supervised learning of universal sentence representations from natural language inference data," *arXiv preprint arXiv:1705.02364*, 2017.

[24] O. Dušek and Z. Kasner, "Evaluating semantic accuracy of data-to-text generation with natural language inference," in *Proc. of INLG*, 2020.