Leveraging AI to Assist Emotional Supports in Online Mental Health Community

DONGHOON SHIN, Seoul National University, Korea
SUBEEN PARK, Seoul National University, Korea
ESTHER HEHSUN KIM, Seoul National University, Korea
SOOMIN KIM, Seoul National University, Korea
JINWOOK SEO, Seoul National University, Korea
HWAJUNG HONG, KAIST, Korea

Although sharing emotional supports in online mental health communities (OMHCs) is an effective and accessible way of managing mental wellbeing, it is often difficult for both seekers and providers. To support empathetic interactions, we design an AI-infused workflow that allows users to write emotional supporting messages to other users' posts based on the elicitation of the seeker's emotion and contextual keywords from writing. Based on a preliminary user study (N = 10), we identified that the system helped seekers to clarify emotion and describe text concretely while writing a post. Providers could also learn how to react empathetically to the post. Based on these results, we suggest design implications for our proposed system.

1 INTRODUCTION

Online mental health communities (OMHCs) have become a prevalent medium of promoting mental wellness through collaborative interactions among people [24]. For instance, users in multiple mental health sub-Reddits benefit from actively sharing their challenging experiences and gaining feedback from their peers online [24]. By posting their experiences and concerns and reacting to them, individuals in OMHCs can actively share emotional and informational support with people with similar conditions [4].

Here, ensuring active emotional support-sharing processes among users is particularly crucial to maintaining thriving OMHCs [8], yet often considered challenging [22]. For providers, it is often burdensome and overwhelming to be emotionally engaged [2, 11], leading to their dropouts [5]. Plus, since most of the existing OMHCs are text-based [12], providers have difficulty in converting their empathetic thoughts in a text-based form [9, 20]. Similarly, it is equally important, yet difficult, for seekers to ensure that they disclose their experiences concretely to help providers better understand seeker's experiences and react emotionally [1, 25].

In this work, we aim to explore the feasibility of AI in augmenting the overall workflow of empathetic communication by considering both seeker and provider sides in OMHCs. Specifically, we designed and evaluated an AI-infused mental health community app, which supports the scaffolded interactions (writing, exploring, and reacting to posts) specialized to facilitate empathetic communication in OMHCs by detecting the emotion and contextual keywords and recommending appropriate triggers/prompts to assist provider's reaction. From the preliminary user study (N = 10), we identified the feasibility and areas of improvement for enhancing emotional support processes in AI-infused OMHCs.

2 PROTOTYPE DESIGN

2.1 Overall Concept and Workflow

The goal of our system is to support empathetic peer-support workflow in OMHCs. Here, in addition to identifying the *emotion* of the seeker's post, it is also important to ensure that providers understand the overall *contextual information* (e.g., place, reason, subject in conflict) to provide more detailed and empathetic response [21]. Thus, we decided to let users elicit and leverage (i) emotion and (ii) contextual information through their exchange, with the aid of AI.

2.2 System Design

In OMHCs, supports are often shared as a form of *post* and *comment* between seekers and providers. Specifically, we conceptualized this process with the following interactions and designed an iOS application: (i) seekers post their emotional contexts, (ii) providers explore such posts in the community, and they (iii) react to the post [10, 18, 19, 24]:

2.2.1 Writing a post. First, the user is asked to freely describe their personal experiences that required emotional support (Figure 1-(a)). Once the user enters text, the system automatically detects the emotion (①) and contextual keywords (②) from the text, which are considered as key information used for psychological therapy assessment [6]. Specifically, the system recommends (i) one most likely representative emotion and (ii) up to three most salient contextual keywords. In this process, to ensure user agency, the user may also enter their own customized keywords.

For recommending the representative emotion, pre-trained model classifies texts in one among six emotions (*anger*, *sadness*, *happiness*, *surprise*, *fear*, and *distress*). Specifically, this taxonomy follows six basic emotions suggested by Ekman [7], except for ‘disgust’ altered by ‘distress’ to cover as many topics in OMHCs as possible.

2.2.2 Reacting to the post. To assist support providers to react empathetically, we decided to use EPITOME [23], a text-based empathic reaction framework. Specifically, by leveraging emotion detected and saved from the writing process, we decided to support (i) emotional reaction as a form of *trigger*, with the rest of the processes guided by the *prompts* (Figure 1-(c)).

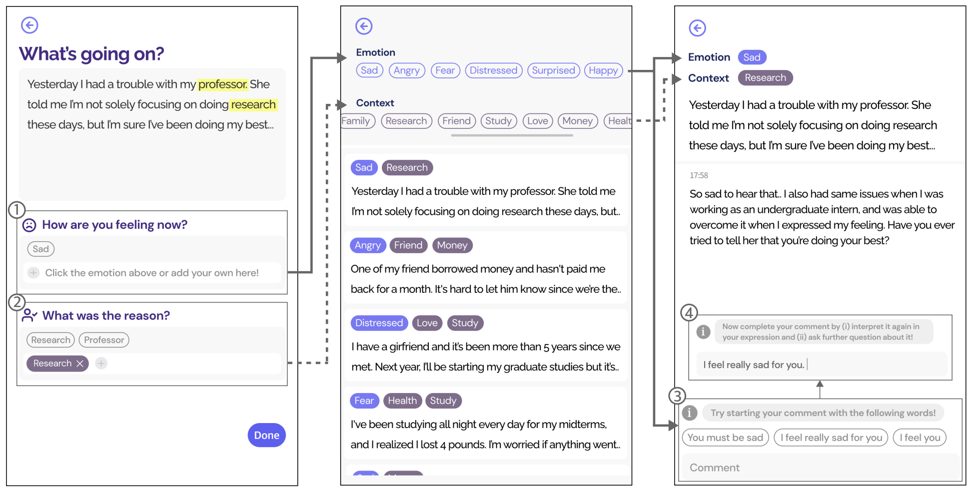
Here, we designed our system to recommend providers with three representative phrases to start their reaction with (trigger; ③). To collect the phrases suitable for users to start with, we first recruited a mental health counselor from the university, who has more than 2 years of counseling experience. Then, we asked them to recommend three short phrases of emotional reaction targeted to each emotion, along with two phrases that can be used for every negative emotion. After gathering all the responses, we designed the system to randomly choose and present two among three targeted phrases based on the detected emotion of the post, and one between two generalized emotions.

Once the user initiates writing a comment by selecting one of the recommended phrases (emotional reaction), they can start entering their own feedback regarding the post based on the prompts offered (④). Here, the system offers prompts that guide users to (i) interpret the situation with their own words (interpretation) and then (ii) throw a question about the post (exploration), consecutively. Each parenthesis above indicates the element of EPITOME [23], the empathic reaction framework we decided to utilize.

2.2.3 Exploring posts. In our interface, filtering tags located on top of the interface can be used to filter posts including specific emotions/contexts. Each tag is the result of detection from each post, and once tags are selected, the system only shows posts containing more than one selected tag until entire tags are deselected.

2.3 Model Selection

For building an emotion classification model, we utilized ELECTRA [3], a state-of-the-art language training model, fine-tuned to the main language of our interface (Korean) [15]. Then, we trained using Korean emotional dialogue corpus [13] (validation set accuracy: 0.7083). Plus, to extract contextual keywords, we used KRWordRank [14], a WordRank-based unsupervised Korean word extraction method. Once the keyword extraction begins, the system triggers the WordRank-based algorithm [17] by default to extract the maximum of 3 keywords from the posting.



(a) Writing a post **(b) Exploring posts** **(c) Reacting to the post**

Fig. 1. Keyscreens of our system we designed and examined. The interface elements were translated from Korean

3 PRELIMINARY USER STUDY

3.1 Methods

3.1.1 Participants. From two major college online communities in South Korea, we recruited participants who self-reported that they often required emotional support and share their emotions in an online medium such as OMHCs or social media. Since we designed and deployed our system targeted to mobile settings, we screened users with the self-reported daily usage of smartphones of at least 4 hours on average for ensuring their familiarity with mobile settings. In addition, people who are currently in the process of mental therapy were excluded to avoid conflict with it. As such, 10 participants were recruited (M_{age} = 22.6, SD_{age} = 1.84; 6 female).

3.1.2 Study procedure. To compare our system with the existing mental health community settings, we designed a *control* interface, whose interface is identical except for the recommendation functionalities offered while users write, explore, and comment on the posts. Then, we ran a within-subject design study where 5 participants were assigned to use *control interface* and *our interface* consecutively, and reversed order for the rest of participants.

First, we briefly introduced the research objectives/process, and the participants installed our prototype application. In each condition, the participants were requested to (i) write two new posts about anxious situations they often faced in daily life, and (ii) write reactions to two of others' posts. For users to follow (ii), we showed participants with pre-defined posts by crawling existing posts from Korean OMHC after removing all the sensitive information.

Each participant was then asked to complete the survey (5-point Likert scale; 1 = strongly disagree, 5 = strongly agree) regarding (i) the ease of using the interface for both our and control system and (ii) satisfaction with the assist of AI for each interaction. Then, we ran a semi-structured interview session to collect their overall experiences of using our system and its future enhancements. Each interview response was transcribed and later open-coded.

Once the study procedures for every participant were complete, each of the participants was compensated with 15,000 KRW (~12 USD). Then, we passed their response data to university mental health counselor and had them evaluate posts and comments based on the following metric on a 5-point Likert scale, respectively: *(i) how much does the post draw emotionally supportive reaction? (ii) how much does the comment show empathy toward the post?* In this process, to ensure the user privacy, we formulated and followed the following precautionary steps: (i) we had our study approved by the university IRB and strictly followed the approved procedure, (ii) all the privacy-sensitive information (e.g., name, affiliation) in transcription and user-provided data was masked once collected.

3.2 Results

3.2.1 Writing a post. Our system induces users to view the detected emotion and contextual keywords and lets them iterate writing based on them. Regarding such an approach, participants first reported that it helped them clarify emotion and circumstance they faced. Plus, our system was reported to help express their thoughts as a form of text concretely, along with assisting users to rely emotionally toward the system.

At the same time, after using our system, user reported the increased awareness of others and reluctance toward a feeling of being 'diagnosed' by AI, which necessitates the design iteration. Plus, analysis from the mental health counselor shows that the self-reported concreteness while explaining their experience was not directed to the increase of the level of inducing empathetic reaction ($t = -0.1787$, $p = 0.43$). Thus, further design consideration is required to lead the concreteness of the post to inducing an empathetic reaction.

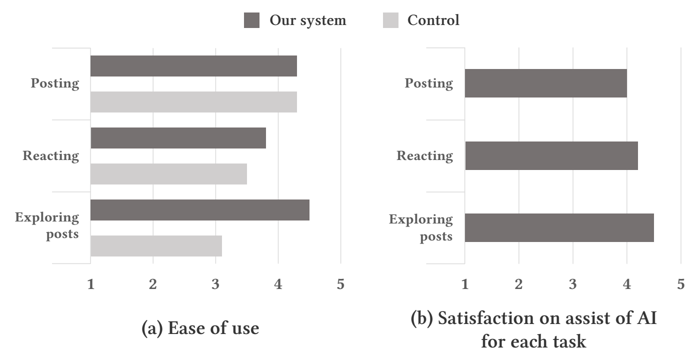


Fig. 2. Survey results of our study

3.2.2 Reacting to the post. From the analysis of mental health counselor, comments from our system presented significant improvement in terms of empathy compared to comments from control ($t = -4.4316$, $p < 0.0001$). We describe several possible attributes that might have led to such enhancement: first, users reported they gained opportunity to learn how to empathize with others. In addition, from the users, we gained feasibility of triggers/prompts in our interface for fostering a healthy community, along with helping users to better understand posts.

However, users also faced several challenges regarding the reaction process. For instance, our triggers/prompts were reported to limit the style of overall supports. Plus, by following the guidance of system for writing the comment, users felt increased burden of comment-writing process.

3.2.3 Exploring posts. Participants showed unanimity on the efficiency of the filtering system in terms of exploring posts. Considering that the tag-based filtering system is commonly found in online communities, such positive feedback implies the feasibility of integrating our AI-driven keyword suggestions into the existing filtering system in OMHCs.

3.3 Discussion

Throughout the study, we could identify the feasibility of AI-assisted writing processes in supporting interpersonal interactions in OMHCs, as well as its feasibility of forming a healthier community environment. Specifically, leveraging

AI-driven emotion/contextual keyword elicitation was reported to induce seekers to clarify expression for AI to better understand, ultimately assisting the writing process to be more concrete. However, from the expert analysis, we realized that such concreteness did not necessarily lead to emotional support. Thus, further design iteration to connect from such concreteness to inducing empathetic reaction would be required.

Plus, our system was reported to assist the support providers by assisting them to learn how to react empathetically. Considering that such an empathetic reaction is a key to thriving OMHCs [16], we could see the feasibility of extending our work to long-term deployment settings. However, considering that some participants worried if our community setting might only be filled with monotonous, emotionally empathetic reactions, it is would also be beneficial to diversify the type of reactions (e.g., supporting energetic reactions) to enrich the community environment.

Still, our study presents several limitations. First, we conducted our study only with 10 participants. Thus, additional participants may be required in terms of generalizability. Second, this study was run in a lab-based setting, where participants were asked to follow designated actions. To collect a more lively experience of users in OMHCs aided by the AI-assisted writing process, a deployment study in the wild might be needed.

REFERENCES

[1] Irwin Altman and Dalmas A Taylor. 1973. *Social penetration: The development of interpersonal relationships.* Holt, Rinehart & Winston.

[2] Lisa J Barney, Kathleen M Griffiths, Helen Christensen, and Anthony F Jorm. 2009. Exploring the nature of stigmatising beliefs about depression and help-seeking: implications for reducing stigma. *BMC public health* 9, 1 (2009), 1–11. https://doi.org/10.1186/1471-2458-9-61

[3] Kevin Clark, Minh-Thang Luong, Quoc V Le, and Christopher D Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555* (2020).

[4] Carolyn E Cutrona and Julie A Suhr. 1994. Social support communication in the context of marriage: an analysis of couples' supportive interactions. (1994).

[5] Fernando Estrada Martinez de Alva, Greg Wadley, and Reeva Lederman. 2015. It feels different from real life: Users' Opinions of Mobile Applications for Mental Health. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction.* 598–602. https://doi.org/10.1145/2838739.2838806

[6] Jill T Ehrenreich, Christopher P Fairholme, Brian A Buzzella, Kristen K Ellard, and David H Barlow. 2007. The role of emotion in psychological therapy. (2007). https://doi.org/10.1111/j.1468-2850.2007.00102.x

[7] Paul Ekman. 1999. Basic emotions. *Handbook of cognition and emotion* 98, 45-60 (1999), 16.

[8] Robert Elliott, Arthur C Bohart, Jeanne C Watson, and Leslie S Greenberg. 2011. Empathy. *Psychotherapy* 48, 1 (2011), 43. https://doi.org/10.1037/a0022187

[9] James Gibson, Doğan Can, Bo Xiao, Zac E. Imel, David C. Atkins, Panayiotis Georgiou, and Shrikanth S. Narayanan. 2016. A Deep Learning Approach to Modeling Empathy in Addiction Counseling. In *Proc. Interspeech 2016.* 1447–1451. https://doi.org/10.21437/Interspeech.2016-554

[10] Xinning Gui, Yu Chen, Yubo Kou, Katie Pine, and Yunan Chen. 2017. Investigating Support Seeking from Peers for Pregnancy in Online Health Communities. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–19. https://doi.org/10.1145/3134685

[11] Nicole J Highet, Bernard G McNair, Tracey A Davenport, and Ian B Hickie. 2004. "How much more can we lose?": carer and family perspectives on living with a person with depression. *Medical Journal of Australia* 181 (2004), S6–S9. https://doi.org/10.5694/j.1326-5377.2004.tb06347.x

[12] Simon Hoermann, Kathryn L McCabe, David N Milne, and Rafael A Calvo. 2017. Application of Synchronous Text-Based Dialogue Systems in Mental Health Interventions: Systematic Review. *Journal of Medical Internet Research* 19, 8 (2017), e7023. https://doi.org/10.2196/jmir.7023

[13] AI Hub. 2020. Korean emotional dialogue corpus. https://aihub.or.kr/aidata/7978

[14] Hyun-joong Kim, Sungzoon Cho, and Pilsung Kang. 2014. KR-WordRank: An Unsupervised Korean Word Extraction Method Based on WordRank. *Journal of Korean Institute of Industrial Engineers* 40, 1 (2014), 18–33. https://doi.org/10.7232/JKIIE.2014.40.1.018

[15] Kiyoung Kim. 2020. Pretrained Language Models For Korean. https://github.com/kiyoungkim1/LMkor.

[16] Taisa Kushner and Amit Sharma. 2020. Bursts of activity: Temporal patterns of help-seeking and support in online mental health forums. In *Proceedings of The Web Conference 2020.* 2906–2912. https://doi.org/10.1145/3366423.3380056

[17] Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing Order into Text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing.* 404–411. https://aclanthology.org/W04-3252

[18] Zhenhui Peng, Qingyu Guo, Ka Wing Tsang, and Xiaojuan Ma. 2020. Exploring the Effects of Technological Writing Assistance for Support Providers in Online Mental Health Community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–15. https://doi.org/10.1145/3313831.3376695

[19] Zhenhui Peng, Xiaojuan Ma, Diyi Yang, Ka Wing Tsang, and Qingyu Guo. 2021. Effects of Support-Seekers' Community Knowledge on Their Expressed Satisfaction with the Received Comments in Mental Health Communities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3411764.3445446

[20] Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Satinder Singh, and Lawrence An. 2017. Understanding and predicting empathic behavior in counseling therapy. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1426–1435. https://doi.org/10.18653/v1/P17-1131

[21] Ulrike Pfeil and Panayiotis Zaphiris. 2007. Patterns of empathy in online communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 919–928. https://doi.org/10.1145/1240624.1240763

[22] Julie Prescott, Terry Hanley, and Katalin Ujhelyi Gomez. 2019. Why do young people use online forums for mental health and emotional support? Benefits and challenges. *British Journal of Guidance & Counselling* 47, 3 (2019), 317–327. https://doi.org/10.1080/03069885.2019.1619169

[23] Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. 5263–5276. https://doi.org/10.18653/v1/2020.emnlp-main.425

[24] Eva Sharma and Munmun De Choudhury. 2018. Mental Health Support and its Relationship to Linguistic Accommodation in Online Communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3173574.3174215

[25] Diyi Yang, Zheng Yao, Joseph Seering, and Robert Kraut. 2019. The Channel Matters: Self-disclosure, Reciprocity and Social Support in Online Cancer Support Groups. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15. https://doi.org/10.1145/3290605.3300261