MindfulDiary: Harnessing Large Language Model to Support Psychiatric Patients' Journaling

Taewan Kim*
KAIST
Republic of Korea
taewan@kaist.ac.kr

Seolyeong Bae†
GIST
Republic of Korea
peixueying@gmail.com

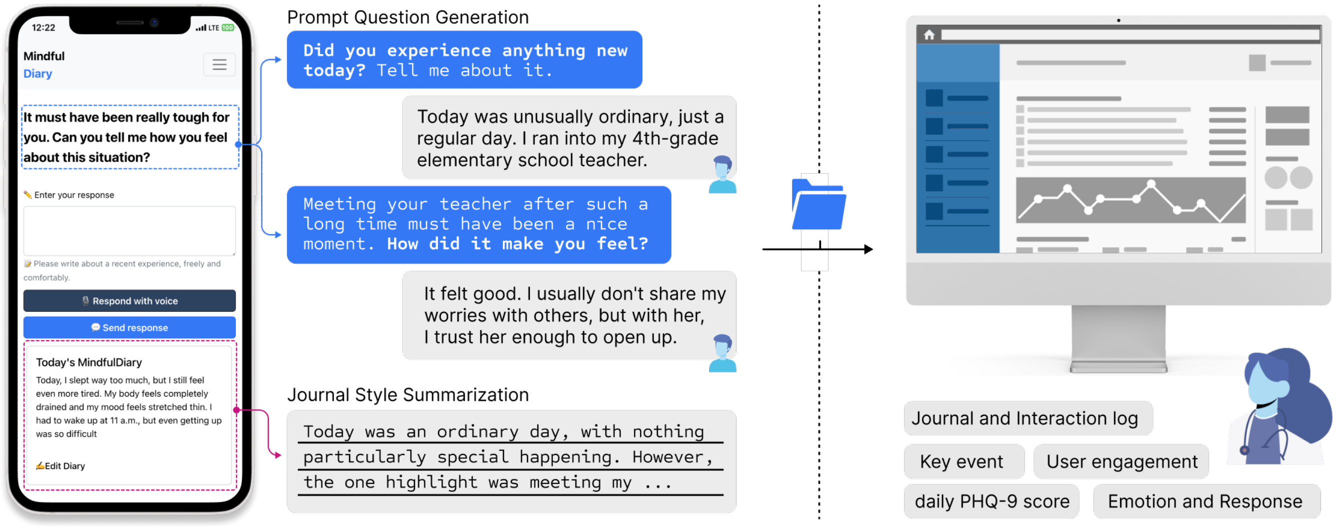
Hyun Ah Kim
NAVER Cloud
Republic of Korea
hyunah.kim@navercorp.com

Su-woo Lee
Wonkwang Univ. Hospital
Republic of Korea
aiesw@naver.com

Hwajung Hong
KAIST
Republic of Korea
hwajung@kaist.ac.kr

Chanmo Yang‡
Wonkwang Univ. Hospital,
Wonkwang University
Republic of Korea
ychanmo@wku.ac.kr

Young-Ho Kim‡
NAVER AI Lab
Republic of Korea
ygho.kim@navercorp.com



It must have been really tough for you. Can you tell me how you feel about this situation?

✏️ Enter your response

📝 Please write about a recent experience, freely and comfortably.

🎙 Respond with voice

💬 Send response

Today's MindfulDiary

Today, I slept way too much, but I still feel even more tired. My body feels completely drained and my mood feels stretched thin. I had to wake up at 11 a.m., but even getting up was so difficult

✍Edit Diary

Did you experience anything new today? Tell me about it.

Today was unusually ordinary, just a regular day. I ran into my 4th-grade elementary school teacher.

Meeting your teacher after such a long time must have been a nice moment. How did it make you feel?

It felt good. I usually don't share my worries with others, but with her, I trust her enough to open up.

Today was an ordinary day, with nothing particularly special happening. However, the one highlight was meeting my ...

(1) Patient Interface

(2) Clinician Dashboard

Figure 1: MindfulDiary consists of two components: the patient interface and the expert interface: (1) Patient interface aids users in daily journaling, offering prompt questions through conversations and crafting summaries in a diary-like format. (2) Clinician Dashboard features a dashboard that visualizes data from patient records, facilitating its use during consultations. (Refer to our supplementary video for the user interaction.)

ABSTRACT

Large Language Models (LLMs) offer promising opportunities in mental health domains, although their inherent complexity and low controllability elicit concern regarding their applicability in clinical settings. We present MindfulDiary, an LLM-driven journaling app that helps psychiatric patients document daily experiences through conversation. Designed in collaboration with mental health professionals, MindfulDiary takes a state-based approach to safely comply with the experts' guidelines while carrying on free-form conversations. Through a four-week field study involving 28 patients with major depressive disorder and five psychiatrists, we examined how MindfulDiary facilitates patients' journaling practice and clinical care. The study revealed that MindfulDiary supported patients in consistently enriching their daily records and helped clinicians better empathize with their patients through an understanding of their thoughts and daily contexts. Drawing on these findings, we discuss the implications of leveraging LLMs in the mental health

---

*Taewan Kim conducted this work as a research intern at NAVER AI Lab.
†Seolyeong Bae conducted this work as an engineering intern at NAVER Cloud.
‡Co-corresponding authors.

This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '24, May 11–16, 2024, Honolulu, HI, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642937

domain, bridging the technical feasibility and their integration into clinical settings.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI; Natural language interfaces**.

KEYWORDS

journaling, chatbot, mental health, clinical setting, psychiatric patient, large language models

ACM Reference Format:
Taewan Kim, Seolyeong Bae, Hyun Ah Kim, Su-woo Lee, Hwajung Hong, Chanmo Yang, and Young-Ho Kim. 2024. MindfulDiary: Harnessing Large Language Model to Support Psychiatric Patients' Journaling. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (*CHI '24*), *May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3613904.3642937

1 INTRODUCTION

Journals serve as a written record of an individual's past events, thoughts, and feelings, allowing genuine expression [89, 90]. Journaling helps people describe experiences and express emotions related to both negative [70, 71] and positive experiences (*e.g.*, growth potential) [9, 35], thereby reducing stress, anxiety, and depression. Prior work has shown the advantages of journaling in clinical mental health contexts, as journals frequently capture patients' daily experiences, symptoms, and other contextual data that are challenging to gather during brief hospital visits [27, 100]. Furthermore, these patient journals can enhance mental health professionals (MHPs) comprehension of their patient's conditions, leading to improved treatment quality [95]. However, writing about one's past feelings and thoughts can be a complex process because people differ in their ability to understand, identify, and verbalize their emotions [78]. In addition, patient under psychotherapy struggle with constructing a narrative and understanding their past [23, 72].

Conversational AIs, or chatbots, have the potential as an alternative form of journaling, easing the collection of personal data. Researchers in the field of Human-Computer Interaction (HCI) have shown that chatbots can help individuals articulate and share their daily experiences. For instance, chatbots to elicit people's self-disclosure can ease the process of emotional expression by providing a safe and supportive environment for individuals to share their experiences and emotions [16, 52, 53, 68]. Furthermore, a machine's inherent trait of not showing fatigue can make people more confident to share their stories truthfully and comfortably [44, 68]. However, existing chatbot prototypes have commonly employed rule-based or retrieval-driven approaches [1], which have limited capability of generating versatile responses following up serendipitous topics during conversation [38, 41, 51]. This trend presents missed opportunities and a lack of understanding regarding conversational AIs that assist with journaling by suggesting, questioning, and empathizing based on the user's diverse experiences.

The recent achievement of Natural Language Processing in large language models (LLMs) opened up new opportunities for bootstrapping chatbots that can carry on more naturalistic conversation [8, 14, 41, 77, 94]. Their capabilities accelerated the development of chatbots in varied topics that can benefit from open-ended conversation, such as regular check-up calls [8, 41], personal health tracking [94], and personal events and emotions [80]. Despite such opportunities, LLMs' inherent uncertainty in control of response generation calls for precautions to handle unintended or inaccurate responses [26, 41, 47, 93]. If applied to clinical and mental health domains, LLM's behaviors should be designed in collaboration with domain experts regarding the relevance and safety of responses.

In this work, we present a case of collaborative design, development, and evaluation of an LLM-infused conversational AI system designed to facilitate the self-reflection of patients and communication with MHPs. We designed and developed **MindfulDiary** (Figure 1), which consists of (1) a mobile conversational AI with which patients can converse about daily experiences and thoughts and (2) a web dashboard that allows MHPs to review their patients' dialogue history with the AI. MindfulDiary incorporates LLMs to generate a response, prompting patients differently according to the conversational phase. The conversation records are automatically summarized and presented on a clinician dashboard so MHPs can obtain insights about the patient.

As a multi-disciplinary research team, which included HCI researchers, AI engineers, and psychiatrists, we iteratively designed MindfulDiary and conducted a four-week field study involving 28 psychiatric patients diagnosed with major depressive disorder (MDD) and five psychiatrists who care for them. During the study, the patients freely used MindfulDiary to record daily conversations, and the psychiatrists used the clinician dashboard during regular clinical visits. Through this study, we found that the versatility, narrative-building capability, and diverse perspectives provided by MindfulDiary assisted patients in consistently enriching their daily records. Furthermore, MindfulDiary supported patients in overcoming the challenges of detailed record-keeping and expression, often hindered by feelings of apathy and cognitive burdens. The psychiatrists reported that enhanced records provided by MindfulDiary offered a more nuanced understanding of their patients, fostering empathy. In addition, MindfulDiary supplemented their consultation by eliciting candid thoughts from patients that may be invasive to be asked by the MHPs.

The key contributions of this work are:

(1) Design and development of MindfulDiary, an LLM-driven journal designed to document psychiatric patients' daily experiences through naturalistic conversations, designed in collaboration with MHPs.

(2) Empirical findings from a four-week field study involving 28 patients and five psychiatrists, demonstrating how MindfulDiary supported patients in keeping their daily logs and assisted psychiatrists in monitoring and comprehending patient states. We also explore how MindfulDiary enhances the quality of patient-provider communication, emphasizing the role of LLMs in prompting deeper self-exploration, which can be instrumental in clinical settings.

(3) Implications for designing and instrumenting LLM-infused conversational AIs in clinical mental health settings.

2 RELATED WORK

In this section, we cover related work in three parts: (1) Journaling for self-reflection and mental Health, (2) Journaling as patient-generated health data in clinical settings, and (3) Conversational agents for mental health.

2.1 Journaling for Self-Reflection and Mental Health

Journaling—recording various personal matters ranging from observations and travels to overall daily experiences and thoughts [76]—takes place in various forms, including letter-style entries reminiscent of Anne Frank's diary or more traditional prose. By narrating personal experiences and innermost feelings in a journal, people discover insights upon reflecting on their past events [89, 90], thereby improving mental wellness and quality of life [9, 70, 71, 91]. Researchers have found that journaling impacts mental health by fostering emotional inhibition [69], cognitive processing [36], and "freeing up" cognitive load [45, 74].

Beyond traditional pen and paper, research in HCI has explored how technology can augment a journaling method in articulating past emotions and experiences (*e.g.*, [4, 10, 24, 37, 37, 42, 46]). Early studies in journaling focused on how technology can assist users in better documenting past events and experiences from life-logging perspectives [24, 37, 42]. For example, SenseCam has been proposed as a wearable ubiquitous computing device that utilizes camera sensor data to not only record but also reflect on the wearer's daily life [37]. Subsequently, various attempts have been made to utilize contextual data and cues such as activity levels [46], mood [10], location, and photos [4] in journaling context.

Furthermore, HCI researchers have investigated ways to improve a journaling method to encourage users to express themselves in a manner that is more comfortable [34], engaging [58], and honest [68]. For example, Park *et al.* found that conversational agents create a social-like environment, encouraging self-reflection and enhancing expressive writing [68]. Furthermore, social support from online communities has been found to increase user participation in journaling activities [58]. Gonzales *et al.* [34] proposed an approach to mitigate the discomfort of revisiting negative memories using the sound generated from the data for more pleasurable [34].

Our work extends this line of research to enhance the journaling experience with technology. Specifically, we explore the potential of interactive dialogue with LLMs as a new format for journaling, allowing psychiatric patients to explore and reflect on their past experiences and emotions. To this end, we leverage a conversational agent as a complementary tool for assisting patients by providing prompt questions to engage users in deeper and more detailed documentation and reflection [46]. We particularly investigate how our approach plays a role in clinical settings as patient-generated health data, which we will cover in the following.

2.2 Journaling as Patient-Generated Health Data in Clinical Settings

Patient-Generated Health Data (PGHD)—defined as "*health-related data, such as health history, symptoms, biometric readings, treatment history, lifestyle choices, and other pertinent details, created, recorded, or gathered by patients*" [81]—has increasingly become an essential tool in clinical settings to capture authentic, real-time insights into patients' health. Studies have shown that PGHD can enhance communication between patients and MHPs and offer contextual information about patients, thereby heightening MHPs' awareness of patient health outside regular clinical visits [20, 55, 64]. For instance, photo journaling improved patient-provider communication for the management and treatment of irritable bowel syndrome by facilitating more effective discussions during treatment [18].

Within the mental health domain, PGHD range from structured mental health assessments (*e.g.*, anxiety, depression) to more unstructured data, including mood-related symptoms and social interactions tracking (*e.g.*, social media use, number of calls)[6, 21, 63]. Patient diaries or journals–"*instructing the patient to write down one's symptoms and other information related to one's daily life to discuss them during clinical appointments*" [27, 100]–can particularly be useful in mental health contexts as it can offer rich, self-documented insights, which could improve MHPs' understanding of their patients [95]. Despite the benefits of patient journaling, people often struggle with starting their entries, sticking to consistent journaling routines, and structuring their reflections [89]. Further, writing about emotions and past experiences can be intricate, as individuals vary in their capacity to recognize, interpret, and articulate their feelings [49, 57, 78]. For some, especially in psychotherapy, crafting a narrative that describes one's life journey can be a challenging process [23, 72].

In this work, we aim to lower the barrier of journaling for patients by allowing them to carry on casual conversations with an AI instead of plain open-ended text writing. We demonstrate the potential of these dialogues as a source of PGHD in clinical settings to facilitate individual self-reflection and enhance communication between patients and MHPs.

2.3 Conversational Agents for Mental Health

The field of AI has proposed significant innovations in medical settings, such as aiding clinical decision-making and diagnosis [11, 60]. In a mental health domain, Natural Language Processing (NLP) techniques have been widely applied to venues that demand human language comprehension and generation for patient support and treatment [15, 22]. Since caring mental health often involves counseling, conversational agents, also known as chatbots, have particularly stood out in the mental health domain [1, 3]. Studies have demonstrated the potential of chatbots in facilitating different types of therapy, such as cognitive behavioral therapy [28, 29, 39], expressive writing [68], behavioral reinforcement [39], and solution-focused therapy [29]. Prior work has also shown that chatbots could ease the burden of disclosing sensitive information. Studies indicated that individuals may feel more comfortable communicating with chatbots because of the social stigma involved in communicating with human beings [52, 53, 68]. Furthermore, these approaches can help overcome temporal and spatial constraints, offering mental health support that is accessible at any time and anywhere [16].

Early mental health therapy chatbots predominantly relied upon rule-based or retrieval-based approaches [1, 2]. While these approaches provide a high level of control over the conversational flow, they fall short of carrying on open-ended conversations; that is, the chatbots neither respond to serendipitous topics nor offer versatile responses that are outside the scope of design [31, 38, 51]. In the mental health context, these limitations may impact the quality of the chatbot's caring behaviors as their messages tend to be general [38, 68]. Recent LLMs have suggested a new paradigm of bootstrapping chatbots [93, 94]. LLM-driven chatbots tend to produce human-like, context-aware responses for unseen topics [65, 94]. Consequently, LLM-driven chatbots excel in facilitating open-domain dialogues and engaging in unscripted interactions [8], offering flexibility and adaptability in conversations [73, 86], especially in complex scenarios like mental health support. For instance, GPT-3.5 demonstrates empathetic traits, such as recognizing emotions and offering emotionally supportive replies in various situations, predominantly in healthcare settings [84]. In certain cases, these models appear to have potential in tasks that involve empathy, showing promising results when compared to humans [7, 25]. Given these early findings, LLM-driven chatbots in the public health sector may offer some support in alleviating emotional burden and loneliness among isolated individuals, though this area is still under exploration [41]. Additionally, the integration of LLMs with human efforts in creating mental health peer support messages could potentially lead to more empathetic conversations [82].

However, LLMs suffer from inherent challenges tied to their transformer-based architecture [92]. One key issue is the explainability of the model output: it is challenging to discern how this 'black box' model interprets a given input prompt. As a result, designers struggle to predict both the LLM's understanding of the input and the subsequent messages it might produce [54]. For instance, a chatbot leveraging GPT-2 for mental health therapy occasionally generated non-word sentences or produced more negative outputs than positive ones [93]. Replika, an LLM-driven application intended for mental well-being, has occasionally displayed harmful content and exhibited inconsistent conversational styles, undermining its role as a long-term companion [59].

These findings highlight two implications for LLM-infused systems in clinical settings. First, human oversight and moderation are critical when using LLMs in clinical settings [93]. To ensure safe and ethical instrumentation of LLMs for mental health patients, this work involves MHPs to reflect domain experts' perspectives [88] in developing an LLM-driven chatbot system. Second, it is necessary to enhance the controllability of an LLM [41, 97] to better follow the intended conversational design. As one exemplar approach to enhance LLM's controllability, yet not a conversational domain, AI Chains break down a complex task into multiple, simpler sub-tasks so that individual LLM inferences can work in better reliability [96]. Similarly, in our system, the chatbot operates using state-based prompting, where a model prompt contains the instruction focused solely on the current state, which is part of an overarching conversational protocol. By simplifying the model prompts, we intended the generated responses to safely and reliably comply with the MHP's guidelines for interacting with the patients.

In summary, our work leverages LLMs in two key components: First, we use an LLM to power a chatbot for patient journaling. Second, the clinician dashboard incorporates LLMs for various NLP tasks, such as text summarization and classification, to visualize summarized insights (*e.g.*, [5, 32]) that are noteworthy for MHP's treatment. Through this collaborative approach involving clinicians, our work explores a unique design space for designing deploying LLM-driven chatbots in the mental health domain, aiming to improve communication between patients and providers by facilitating the recording of daily experiences, which act as a bridge for a better understanding of patients.

3 FORMATIVE STUDY: FOCUS GROUP INTERVIEW

To inform the design of MindfulDiary, we first conducted a Focus Group Interview (FGI) with MHPs. The goal of the FGI was to understand MHPs' perspectives, expectations, values, and precautions in utilizing LLMs in the clinical mental health context. Based on this understanding, we aimed to design the functions and interactions that the system should provide. This was an essential process in our overall approach, not just technology-centered system design, but creating a system meaningful to users and stakeholders [87].

3.1 Procedure and Analysis

We distributed recruitment flyers in the Department of Psychiatry at a local university hospital, inviting Mental Health Professionals (MHPs) working in departments of psychiatry and mental health care centers to participate. We recruited six MHPs (E1–6; two males and four females)—four clinical psychologists and two psychiatrists whose careers varied from 1 to 11 years. Four were clinical psychologists responsible for counseling and daily monitoring and intervention of at-risk patient groups in local mental health centers and university hospitals, and two were psychiatrists in charge of outpatient and inpatient ward treatment in the psychiatry department of university hospitals (see Table 1).

We invited participants to two 1-hour remote sessions on Zoom. Two researchers participated in the sessions. We first provided an overview of language model technologies and LLM's natural language understanding and generation capabilities until we shared a common understanding of the principles, applications, opportunities, and limitations of LLMs. Considering that we were designing a system for individuals with mental health challenges, we thoroughly covered the drawbacks of LLMs, such as uncertainty in control and hallucinations.

After the overview, we went through group discussions on how LLMs could be utilized in the current patient treatment process. As a probe, we asked participants a focused set of questions on (1) the challenges MHPs currently face during patient treatment and counseling sessions, and (2) their expectations and envisioned opportunities of LLMs' role in clinical mental health settings. We sought to understand the experts' perspectives through questions such as, '*What are the difficulties or challenges patients face in their daily lives between treatments (or counseling)?*', '*What are the important considerations in self-care that patients perform in their daily lives?*', and '*What questions or conversational techniques do you use to encourage patients to share about their daily lives and moods?*'.

Table 1: Demographic of FGI Participants (E1–6).

Code	Gender	Age	Job title	Years of Experience
E1	F	34	Clinical psychologist	2 years
E2	F	30	Clinical psychologist	1 year
E3	F	38	Clinical psychologist	9 years
E4	F	36	Clinical psychologist	11 years
E5	M	35	Psychiatrist	2 years
E6	M	40	Psychiatrist	10 years

The session was video recorded and later transcribed. We open-coded the transcripts to identify emerging themes. In the following, we cover the findings from the FGI.

3.2 Findings from the Interviews

3.2.1 Challenges in Eliciting Responses from Patients with Depression. Participants indicated that eliciting disclosure from patients' inner thoughts during a limited consultation time requires significant effort. Many patients with depression experience difficulty describing and expressing their feelings and thoughts to providers due to a sense of apathy, which is a common psychiatric symptom involved in Major Depressive Disorder: "*In the consultation room, even if they sit like this, they often just remain silent for a long time.*" (E5) Thus, providers often end up spending a substantial amount of time asking standardized and repetitive questions about mood, sleep, and major events to understand patients' current states.

Participants also noted that they had their patients engaged in paper-based diary writing methods but most demonstrated low participation rates and low engagement: "*We tried a diary method on paper(in the inpatient ward), and several patients did write. What we saw was quite trivial, like, 'I just felt bad today.' But we learned there were significant events upon consultation, like having a big argument with other patients, which they did not record. Because patients with depression, or those who have had suicidal or self-harming incidents, often have a dulled state in expressing their emotions or feel apathetic, they tend to find such expressions very difficult.*" (E3)

3.2.2 LLMs as a Bridge for Enhanced Patient Communication. Our participants envisioned LLMs as a transformative tool in mental health care, particularly for enhancing interactions with patients who struggle to express themselves. They recognized that the natural and flexible conversational abilities of LLMs could bridge communication gaps, offering a more nuanced understanding of patients' conditions. This could be particularly beneficial in cases where patients have difficulty articulating their feelings due to symptoms like apathy or social phobia. Additionally, participants noted that using LLMs could be significantly more interactive and engaging than traditional paper-based approaches, potentially increasing compliance and participation in the therapeutic process.

Participants especially underscored the importance of capturing the continuum of thoughts leading up to a particular emotional state, such as fear, in the journaling process. They envisioned the need for using LLMs to introspect deeper into the patient's psyche, revealing underlying thoughts and emotions that the patient might not be consciously aware of. E5 mentioned, "It would be good if the journal continuously records the flow of thoughts. For example, it would be beneficial to document the various thoughts and detailed reasons leading up to certain feelings like fear. Like, 'I feel scared when I'm in a place with many people,' and then digging deeper into 'Why do I feel scared?'—I think a process that gets more specific like this would be good." This approach not only aids in a more comprehensive self-examination but also enriches the therapeutic dialogue between the patient and the MHP.

3.2.3 LLMs for Analytical Insights and Personalized Mental Health Support. The participants further suggested that LLMs could analyze journal entries to identify key themes, words, or sentiments expressed over time, offering patients tangible feedback on their emotional patterns and progress. Such analytical capabilities could empower patients with a greater sense of control and awareness of their mental health journey, potentially motivating them towards self-management and active participation in their treatment. Additionally, the analysis could assist MHPs in a deeper understanding of their patient's emotional states and thought processes by examining the tone, choice of words, and speech or writing patterns. The participants envisioned that insights derived from LLMs about patient journaling habits could inform MHPs about the most effective counseling approaches for each individual. They suggested, *"Observing how patients react to different forms of communication can provide valuable information. Some patients might find solace in simple reassurance, while others may benefit from more straightforward, targeted feedback."*

3.3 Improvements after the Interviews

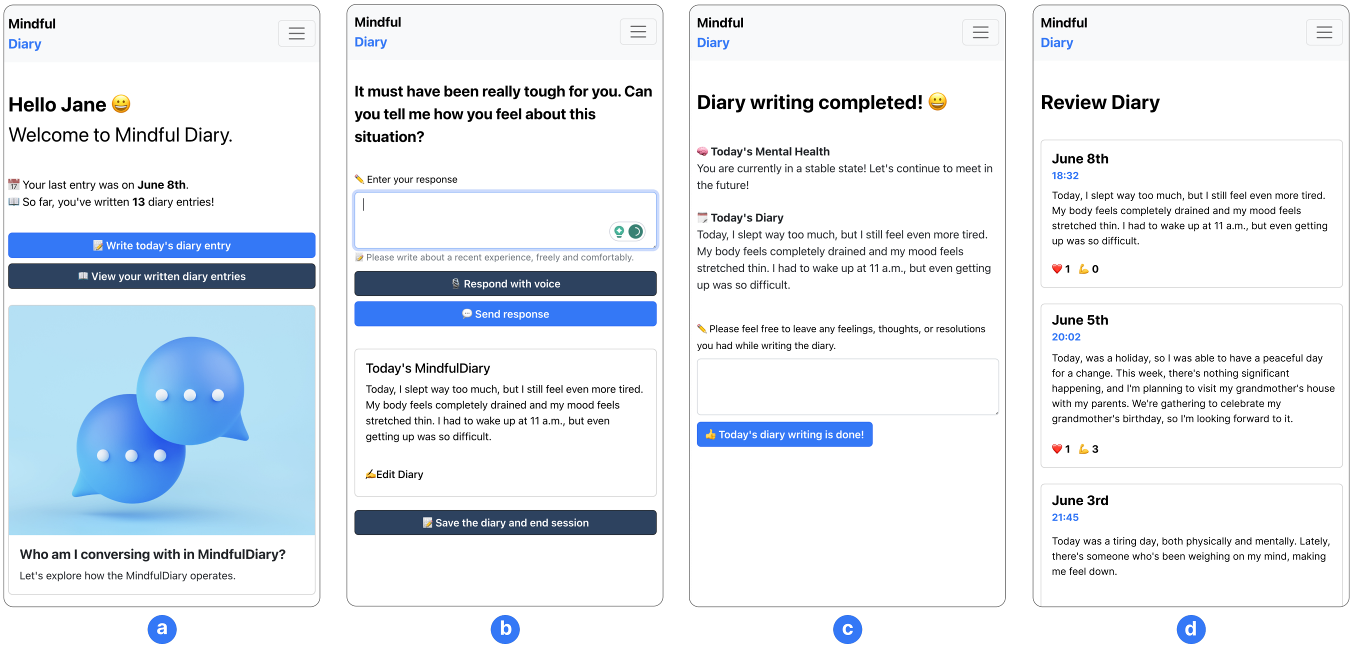
Based on the lessons from the FGI, we refined the initial concept of MindfulDiary. We leveraged the conversational abilities of LLMs to help patients document their daily experiences between clinical visits. MHPs had access to the collected data to inform their clinical decision-making. Furthermore, both MHPs and the research team concur that LLMs should not act solely as the primary intervention due to their inherent limitations but should function as supportive tools for clinical consultations. The subsequent section outlines the design and development process of our system.

4 MINDFULDIARY

Informed by the findings from FGI with MHPs, we designed and developed MindfulDiary, which consists of two main components: (1) a patient mobile app for daily record-keeping and (2) a clinician dashboard that allows professionals to access and use these daily records in a clinical setting (See Figure 1). Below, we present a fictional usage scenario to demonstrate how the system works.

Jane, diagnosed with chronic anxiety, frequently grapples with panic attacks. To keep track of her daily experiences, her psychiatrist recommends trying MindfulDiary as part of her treatment plan.

Every evening, Jane converses with the MindfulDiary app regarding her daily activities, emotions, and thoughts. The AI leads the conversation with Jane by asking prompted and follow-up questions about her day. After a session, the app summarizes the dialogue into a journal-style essay, on which she can revisit and reflect later. She can explore the summarized essays whenever she wants to reflect on past events or thoughts.


Hello Jane 😀

Welcome to Mindful Diary.

📅 Your last entry was on **June 8th**.

📖 So far, you've written **13** diary entries!

Who am I conversing with in MindfulDiary?

Let's explore how the MindfulDiary operates.

It must have been really tough for you. Can you tell me how you feel about this situation?

✏️ Enter your response

📝 Please write about a recent experience, freely and comfortably.

Today's MindfulDiary

Today, I slept way too much, but I still feel even more tired. My body feels completely drained and my mood feels stretched thin. I had to wake up at 11 a.m., but even getting up was so difficult.

Diary writing completed! 😀

🧠 **Today's Mental Health**

You are currently in a stable state! Let's continue to meet in the future!

📆 **Today's Diary**

Today, I slept way too much, but I still feel even more tired. My body feels completely drained and my mood feels stretched thin. I had to wake up at 11 a.m., but even getting up was so difficult.

✏️ Please feel free to leave any feelings, thoughts, or resolutions you had while writing the diary.

Review Diary

June 8th
18:32

Today, I slept way too much, but I still feel even more tired. My body feels completely drained and my mood feels stretched thin. I had to wake up at 11 a.m., but even getting up was so difficult.

❤️ 1 💪 0

June 5th
20:02

Today, was a holiday, so I was able to have a peaceful day for a change. This week, there's nothing significant happening, and I'm planning to visit my grandmother's house with my parents. We're gathering to celebrate my grandmother's birthday, so I'm looking forward to it.

❤️ 1 💪 3

June 3rd
21:45

Today was a tiring day, both physically and mentally. Lately, there's someone who's been weighing on my mind, making me feel down.

Mindful
Diary
📝 Write today's diary entry
📖 View your written diary entries
🎙 Respond with voice
💬 Send response
✍Edit Diary
📝 Save the diary and end session
👍 Today's diary writing is done!
a
b
c
d

Figure 2: Main screens of the MindfulDiary app. (a) The main screen, (b) the journaling screen, (c) the summary screen shown when the user submitted the journal dialogue, and (d) the review screen displaying the user's past journal.

Three weeks later, during a consultation, her psychiatrist uses the expert interface of MindfulDiary to review a data-driven summary of Jane's entries. The data helped the psychiatrist identify patterns that Jane's anxiety often spikes during her work commute. Based on this insight, the psychiatrist refines advice and introduces specific coping strategies, fostering a more personalized approach to care.

4.1 MindfulDiary App

The MindfulDiary app for patients aims to support people who might have difficulty journaling due to apathy and cognitive load through naturalistic conversation driven by an LLM. The app consists of a home screen containing an introduction and guide to the system (Figure 2a), a journal writing screen (Figure 2b, 2c), and a screen to review the diary entries (Figure 2d).

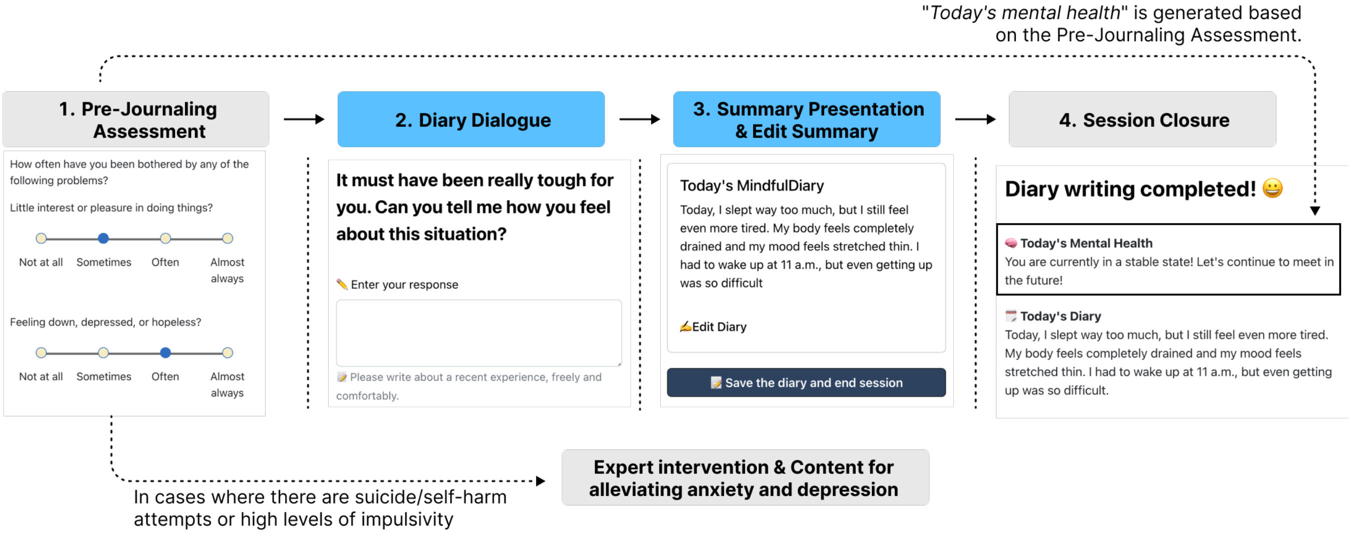
4.1.1 Journaling User Interface. Figure 3 illustrates the overall use flow of the journaling session, which begins with a Pre-Journaling Assessment (Figure 3-①) that asks to fill out a questionnaire for mental health. The questionnaire comprised the modified PHQ-9 [48] and a custom open-ended question inquiring about recent attempts of self-harm or suicide. This assessment prevents users who provided any clues of suicidal or self-harm from journaling on the same day. (We cover this feature in detail in Section 5.4.)

On the next screen, the user converses with MindfulDiary, documenting the events of the day (See Figure 2b). After three turns, MindfulDiary provides a summary of the conversation as an essay. Users can edit this automatically generated summary any time. When the user ends the session by pressing the end button (Figure 2b, bottom), MindfulDiary displays daily mental health insights alongside the diary content on the summary screen (See Figure 2c). Users can also leave a reflection message there. Lastly, users can browse their past records in the Diary Review menu (See Figure 2d).

4.1.2 Conversation Design. We designed the chatbot's conversational behavior based on insights from psychiatry literature [66, 67], which covers foundational techniques and considerations for conducting clinical interviews. We also incorporated the hands-on clinical experiences of practicing psychiatrists.

As a result, we designed the conversation of a journaling session to follow a sequence of three stages: *Rapport building*, *Exploration*, and *Wrap-up*. The **Rapport Building** state is an ice-breaker, centered on casual exchanges about a user's day. In this state, the assistant also shares bits of information to encourage users' openness. This approach is based on previous research findings that a chatbot's self-disclosure positively impacts user disclosure [52] and leverages the natural story-building ability of LLMs [99]. Overall, in this stage, our goal is to create an environment where users can comfortably share their stories. As we progress to the **Exploration** state, the emphasis shifts to a comprehensive understanding of the user's daily events, feelings, and thoughts, facilitated by a mix of open-ended and closed-ended queries that ensure users remain engaged and in control of the dialogue. While open-ended queries are intended to facilitate increase the expression of feelings and emotion and less judgemental, closed-ended queries is for specific and detailed description of the experiences [66, 67].The conversation then transitions to the **Wrap-up**, emphasizing completion and ensuring users have fully voiced their experiences while the system remains empathetic and receptive to any lingering topics.

Besides the three main stages, we also incorporated the **Sensitive Topic** state that handles the most sensitive subjects, such as


"*Today's mental health*" is generated based on the Pre-Journaling Assessment.
1. Pre-Journaling Assessment
2. Diary Dialogue
3. Summary Presentation & Edit Summary
4. Session Closure
Not at all
Sometimes
Often
Almost always
✏️ Enter your response
Today's MindfulDiary
✍️Edit Diary
📝 Save the diary and end session
Diary writing completed! 😀
🧠 Today's Mental Health
📅 Today's Diary
Expert intervention & Content for alleviating anxiety and depression
In cases where there are suicide/self-harm attempts or high levels of impulsivity

How often have you been bothered by any of the following problems?

Little interest or pleasure in doing things?

Feeling down, depressed, or hopeless?

It must have been really tough for you. Can you tell me how you feel about this situation?

📝 Please write about a recent experience, freely and comfortably.

Today, I slept way too much, but I still feel even more tired. My body feels completely drained and my mood feels stretched thin. I had to wake up at 11 a.m., but even getting up was so difficult

You are currently in a stable state! Let's continue to meet in the future!

Today, I slept way too much, but I still feel even more tired. My body feels completely drained and my mood feels stretched thin. I had to wake up at 11 a.m., but even getting up was so difficult.

Figure 3: Use flow of MindfulDiary's journaling session: (1) Pre-Journaling Assessment: Users undergo a mental health survey using the modified PHQ-9 [48] before using MindfulDiary; (2) Users converse with MindfulDiary, documenting their day; (3) Summary Presentation: After three turns, MindfulDiary presents a diary-styled summary of the conversation so far, which can also be edited by the user. Users can continue the conversation as they want. (4) Session Closure: Once all processes are completed, MindfulDiary displays today's mental health and diary content, concluding the journaling session.

self-harm or suicidal ideation. When this state is triggered, psychiatrists receive instant notifications. This allows them to oversee the conversation in real-time and step in to assist the patient if necessary. Here, the system begins by empathizing with the user, recognizing their struggles, and offering a reassuring message. Following this, the system gently probes the depth of their suicidal or self-harm thoughts. If the user expresses intense or specific plans related to self-harm or suicide, the system urges them to seek prompt assistance, either at a hospital or via the local helpline.

4.1.3 Conversational Pipeline. Lengthy and complex input prompts for LLMs are known to cause poor task performance [14] by partly omitting latent concepts [96]. To steer the LLM to comply with the conversational design we intended diligently, we designed MindfulDiary's dialogue system as a state machine. Each conversation stage is carried on with a dedicated input prompt, which is more succinct and clear than a single master prompt containing instructions for all stages.

Figure 4 illustrates our conversation pipeline that runs each time a new user message is received. The pipeline incorporates two LLM-driven modules, a *dialogue analyzer* and the *response generator*.

The **dialogue analyzer** handles the stage transition, returning the stage suggestion—whether to stay or move to a new stage—and a summary paragraph of the current dialogue from the current dialogue information. The dialogue analyzer receives an input prompt that consists of the current number of turns in the conversation (progress level), the most recent stage information, and a list of criteria for recommending each stage (See ② in Figure 4). Based on this information, the underlying LLM performs a summarization task that yields a summarized paragraph of the current dialogue, a recommendation for the next stage based on the summary, turns, and the most recent stage information. For example, the system decides to move to the Wrap-up stage when the user expresses a desire to conclude or say goodbye.

The system then formulates an LLM prompt, combining a dedicated prompt for the current stage, the dialogue summary, and the recent six messages (*i.e.*, three turn pairs) (See ③ in Figure 4). Receiving the prompt as an input, the **response generator** generates an AI message. The stage prompt consists of the description of the task that the LLM is supposed to perform in the current stage, and the speaking rules describing the attitude that the module exhibits in the conversation. For example, the task description of Exploration stage instructs to "`ask questions that encourage users to reflect on their personal stories regarding daily events, thoughts, emotions, challenges, and etc.`" The speaking rules for the Rapport-building stage instruct to keep conversations simple and friendly and reply in an empathetic way.

4.2 Pilot Evaluation

To ensure that MindfulDiary is reliable and safe for conversing with psychiatric patients, we underwent multiple rounds of pilot evaluation. First, we invited five psychiatrists and three clinical psychologists to test the conversational pipeline. The experts provided feedback on the instructions in the model prompts, focusing on their clinical relevance and the embedded terminology and strategies. Then, the experts inspected the chatbot's behavior by chatting with it while role-playing as a patient persona. In particular, we examined the chatbot's reactions to subtle implications of suicide or self-harm in user messages.

After iterating on the conversational pipeline, we conducted a pilot lab study with five patients admitted to a university hospital

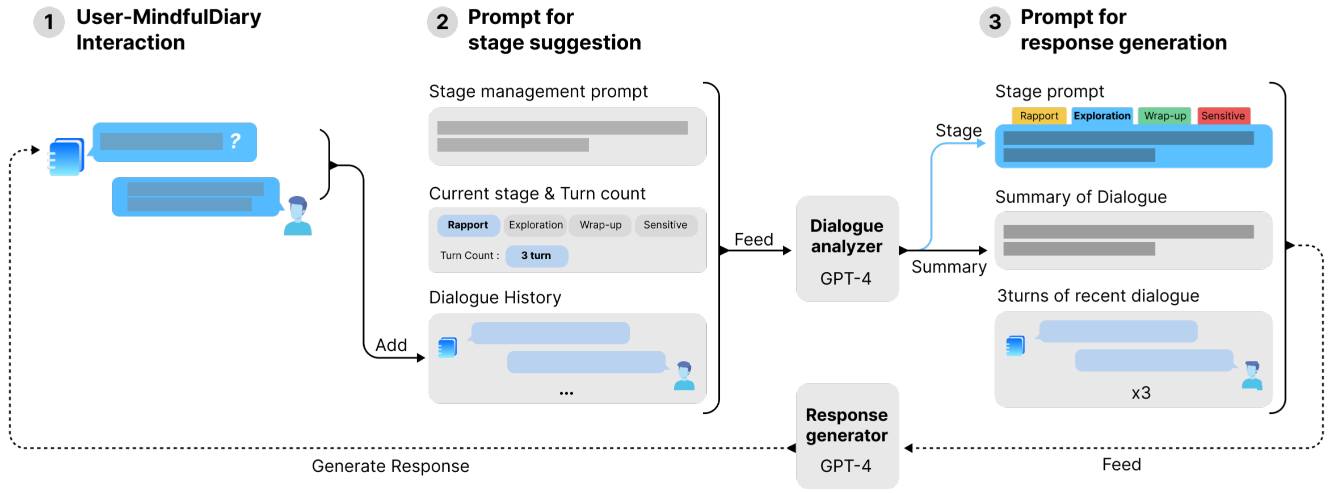


Figure 4: Structure of MindfulDiary's conversational pipeline. (1) Users respond to MindfulDiary's messages. (2) The recent turn count, current state, and whole user-MindfulDiary dialogue are fed into Dialogue Analyzer. (3) Using the output of Dialogue Analyzer, a designated state prompt, a summary of dialogue (containing overall dialogue context), and the latest three conversation turns are fed into Response Generator. The resulting response is then displayed to the user. Both the Dialogue Analyzer and Response Generator operate based on the GPT-4 LLM.

but about to be discharged soon. To ensure safety against risky messages from an LLM, we used a test platform where the participant's clinician monitored the generated messages in real-time, approving them or sending better messages manually.

4.3 Clinician Dashboard

The clinician dashboard (*c.f.*, Supplementary video) is a desktop application designed to facilitate monitoring patient's journal entries and to provide analysis of the entries to help clinicians identify significant events, reactions, and emotions. The dashboard consists of the following components:

User Engagement. This section visualizes the participant's overall engagement with MindfulDiary, including the number of journals written, the date and time they were written, and their length. The modified PHQ-9 scores for each session are also visualized, allowing professionals to track the user's mental health trends using a validated tool.

Journals. This section displays the content of the journals written by patients. The information is presented in a card format, where each card offers a summary of the journal, including timestamps, total time taken to write the journal, and associated PHQ-9 score. The interaction logs between the patient and MindfulDiary are also provided in this section.

Insights. To assist professionals in browsing through the diary, this section visualizes (1) a word cloud to understand frequent terms that the participant used at a glance, (2) a summary of major events to highlight significant happenings and (3) summary of emotions to gauge the mood based on user input. When a specific period is selected for review, a comprehensive summary is generated. We used GPT-4 for most summarization tasks. To generate the word frequency data for the word cloud, we combined GPT-4 and a Korean morphological analysis package named Kiwi [50] to filter only nouns and verbs from the GPT output. Due to the limitations of language model-driven analysis, there might be occasional inaccuracies in the generated content. First-time users of this interface are alerted about possible inaccuracies. An in-interface tooltip also reminds users that the summarized outcomes might not be accurate.

4.4 Technical Implementation

MindfulDiary's interface is developed using React, a JavaScript-based framework. The server, responsible for interfacing with the LLM and overseeing database operations, is implemented in Python. Google Firebase handles user authentication, data storage, and retrieval tasks. The conversational capabilities of MindfulDiary are powered by *gpt-4*, accessible through OpenAI's API[1]. We specifically used `gpt-4-0613` model. For parameter setting, we consistently set the temperature to 0.7 and both a presence penalty and frequency penalty to 0.5.

5 FIELD DEPLOYMENT STUDY

Using MindfulDiary, we conducted a four-week field deployment study with 28 patients undergoing outpatient treatment. Through the study, we aimed to explore how patients and MHPs utilize MindfulDiary and what opportunities and challenges arise from its real-world use. The study protocol was approved by the Institutional Review Board of a university hospital.

5.1 Recruitment

We targeted outpatients from the Department of Mental Health at a University Hospital. Participants were selected based on specific

---

[1] https://platform.openai.com/docs/guides/gpt/chat-completions-api

criteria: (1) those who had been diagnosed with MDD and (2) those who did not exhibit heightened impulsive tendencies or harbor specific intentions towards self-harm or suicide. Key exclusion criteria included a history of psychotic disorders, substance-related disorders, neurodevelopmental disorders, and neurological disorders. Eligible participants were identified through evaluations conducted by psychiatrists. Flyers and consent forms were distributed to eligible patients. For minors, the consent form process was adhered to only when they were accompanied by a guardian at the hospital.

We compensated participants on a weekly basis of participation: For participating every seven days from the starting date, participants received 15,000 KRW (approx. 11 USD). If they completed the entire four-week study process, they received 20,000 KRW as a bonus (*i.e.*, 80,000 KRW—approx. 60 USD—in total). We did not tie the number of dialogue entries to the compensation to ensure natural data entry behavior.

Table 2: Demographic of MindfulDiary participants (P1–28). The table presents gender, age, and the severity of depressive symptoms represented by CES-DC (Center for Epidemiologic Studies Depression Scale for Children), which assesses symptoms of depression in children and adolescents. Scores were categorized as follows: below 16 as Minimal, 16 and above as Mild, and 25 or higher as Severe [17].

Alias	Gender	Age	Severity of depressive symptoms (CES-DC)
P1	F	16	Minimal
P2	F	19	Mild
P3	M	17	Minimal
P4	F	17	Severe
P5	F	14	Severe
P6	F	17	Severe
P7	F	16	Minimal
P8	F	14	Mild
P9	M	16	Mild
P10	F	16	Severe
P11	F	19	Severe
P12	M	19	Severe
P13	F	19	Mild
P14	M	18	Severe
P15	M	24	Minimal
P16	F	12	Severe
P17	F	15	Severe
P18	F	17	Mild
P19	F	23	Severe
P20	M	17	Minimal
P21	M	19	Severe
P22	M	17	Minimal
P23	F	28	Severe
P24	M	17	Minimal
P25	M	19	Minimal
P26	F	19	Severe
P27	M	15	Mild
P28	F	14	Severe

As a minimum requirement for study completion, we instructed the participants not to miss four consecutive days without conversing with MindfulDiary. If a participant missed three consecutive days, an experimenter sent a reminder. In cases where participants did not respond to these reminders, their participation in the study was discontinued. This procedure was implemented to ensure active monitoring and communication. Considering that our system is designed for individuals with mental health challenges, it was crucial to maintain contact with participants and ensure their adherence to the study protocol.

Initially, 36 patients started using MindfulDiary. During the deployment, eight dropped out as they did not meet the minimum data collection requirement. These participants were disengaged from MindfulDiary due to the lack of time or decreased interest. As a result, 28 participants (P1–28; 11 males and 17 females) completed the 4-week field study and were included in the analysis. The majority of participants were adolescents and adults, with ages ranging from 12 to 28 years, with an average age of 17.6 (SD = 3.26). Table 2 presents the demographic details and severity of depressive symptoms of the study participants. These scores are derived from psychiatric evaluations conducted within one week before the starting dates.

5.2 Procedure

Figure 5 illustrates the procedure of the field deployment study. All interviews took place remotely on Zoom.

5.2.1 *MindfulDiary App.*

We deployed the MindfulDiary app to our patient participants. The patient protocol consisted of three parts: (1) an introductory session, (2) deployment, and (3) interviews.

Introductory Session. We first invited each participant to a remote introductory session. A researcher went through our study goal, the motivation of the MindfulDiary system, and the overall procedure of the study. We then played a demo video demonstrating how to use the MindfulDiary app. The session took about 15 minutes.

Deployment. The day following the introductory session, participants started using MindfulDiary for four weeks. We instructed participants to engage with the app whenever they have anything noteworthy but encouraged them to use it at the end of the day. We collected all data from their interactions with the MindfulDiary and the raw input content and outputs from the LLM. We asked our participants to fill out online surveys three times, at the beginning of Week 1, after Week 2, and after the deployment, to measure participants' mental health status and their self-help capability in managing their mental health. The surveys utilized the PHQ-9 [48], GAD-7 [85], and Coping Strategies Scale [98]. (The survey results from the scale are outside the scope of this investigation.)

Mid-study and Debriefing Interviews. We conducted two 15-minute interviews, after the second and fourth weeks, with each participant to understand their experiences and learn how they used MindfulDiary on a daily basis. Considering the characteristics of depression patients, who may struggle to focus for long periods of time, the interview session was divided into two shorter sessions.

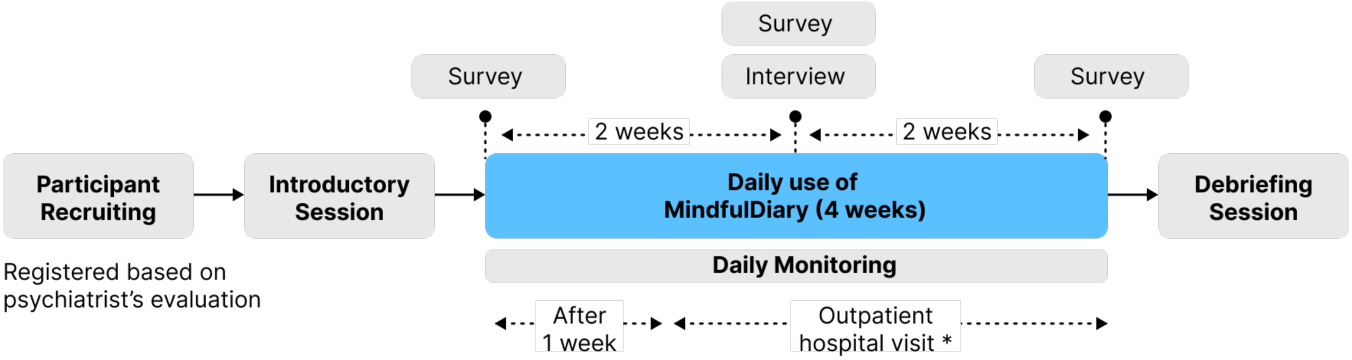


Figure 5: Procedure of the four-week field deployment study: A four-week exploration into the utilization of MindfulDiary by outpatient patients, encompassing daily use, and its integration into clinical decision-making. We note that some participants did not have a follow-up visit during the experimental period. The surveys are outside the scope of this work's investigation.

5.2.2 Clinician Dashboard. Most patient participants had a clinical visit during Week 2 through Week 4 of the deployment period. We deployed MindfulDiary's clinician dashboard to five psychiatrists who are in charge of the participants.

Deployment of Clinician Dashboard in Clinic. We provided instructions to clinicians covering the main components of the clinician dashboard and how to interact with them. To explore the opportunities and limitations of the dashboard, we did not offer explicit instructions for utilizing the clinician dashboard in their workflow. However, we advised psychiatrists to be cautious with the LLM-driven analysis due to potential inaccuracies, emphasizing the importance of verifying data through the interaction logs. The psychiatrists autonomously utilized the clinician dashboard, making sure it didn't disrupt their current treatment methods and preparation routines.

Debriefing Interviews. We interviewed psychiatrists who treated the patient participants to understand how they used the clinician dashboard in clinical settings. We further gathered feedback from the psychiatrists on the opportunities and limitations of MindfulDiary, as well as suggestions for improvements. The interviews with the psychiatrists were conducted offline for about one hour after the deployment study concluded.

5.3 Analysis

To explore participants' usage patterns with MindfulDiary, we first conducted a descriptive statistics analysis. To determine any shifts in participants' adherence over time, we examined weekly writing frequencies using a one-way repeated measures ANOVA (RM-ANOVA) with Greenhouse-Geisser correction. To gain a deeper qualitative understanding of the messages produced by MindfulDiary and interviews with patients and psychiatrists, we used open coding paired with thematic analysis [13]. For a more in-depth qualitative analysis of the messages produced by MindfulDiary and the interviews with patients and psychiatrists, we employed open coding paired with thematic analysis [13]. All interviews were audio-recorded and transcribed for this purpose.

The qualitative analysis was conducted by the first author, a PhD student in HCI, who open-coded the interview transcripts and interaction log data through multiple rounds of iteration. Another author who holds a PhD degree in HCI also contributed to this coding process. Following the initial coding, two psychiatrists reviewed the coded data to provide clinical insights and ensure the accuracy of interpretations. Through discussions among the research team, including these diverse perspectives, overarching themes were identified, enhancing the depth and validity of our qualitative findings.

5.4 Ethical Considerations

Conducting this study, we are fully aware of the inherent risks associated with our research, particularly given the characteristics of participants diagnosed with MDD. To mitigate the risks, we first carefully screened participants, relying on evaluations conducted by psychiatrists. Individuals displaying heightened impulsive tendencies or harboring specific intentions towards self-harm or suicide were excluded from the study. In addition, participants were asked to take the PHQ-9 before interacting with MindfulDiary, along with an additional set of questions probing their recent attempts at self-harm or suicide. If a participant's response to question number 9 of the PHQ-9, regarding suicidal/self-harm thoughts, scored 'moderate or higher' or if any recent suicide attempt was verified, the system pivoted to provide content geared towards alleviating anxiety and reducing stress rather than proceeding with the standard system. In such a case, a real-time alert was also sent to psychiatrists. Lastly, if sensitive themes frequently surfaced in a participant's input during the study, their interactions with the system were temporarily halted. Psychiatrists subsequently re-evaluated such participants to assess the viability of their ongoing participation. During our experiment, for the case of P11, mentions of repetitive suicide and self-harm were detected. Consequently, an expert contacted the participant, the experiment was suspended for three days, and after a re-evaluation in an outpatient clinic, we resumed the system use with P11.

Further, to mitigate potential risks from the LLMs' outputs, we embraced an iterative design methodology. The system's interactions underwent repeated assessments to ensure it generated safe, non-harmful outputs. In addition, in the first week of each participant's system use, all interactions between participants and MindfulDiary were observed in real time. To facilitate this process, when a participant started the session, the research team received a notification email. This notification included real-time monitoring links and reports of the survey responses that participants answered before each session. After the first week, user interactions and MindfulDiary were reviewed within a 12-hour window. During the review process, if an interaction contained sensitive content (specifically, terms pre-defined as sensitive by psychiatrists), the psychiatrists on our research team assessed the situation and contacted the affected participants if necessary.

Lastly, given that we were handling the patients' personal and sensitive data, ensuring the secure protection and management of data was critical. Therefore, during the study, we utilized the Google Firebase authentication service to manage the user authentication process for participants. We were thus able to ensure that only authorized personnel had access to the data, and any attempts at unauthorized access could be promptly detected and managed. After the field study, all data was separated from personal identifiers to maintain anonymity.

6 RESULTS

In this section, we report the results of the field study in four parts: (1) Journaling adherence, (2) Dialogue patterns, (3) Patients' perspectives on MindfulDiary, and (4) MHPs' perspectives of MindfulDiary for clinical settings.

6.1 Journaling Adherence

Figure 6 summarizes the daily engagement of participants with MindfulDiary over the course of four weeks. The colored squares denote the days that participants conversed with MindfulDiary (*i.e.*, days with interaction). Across four weeks, participants submitted 501 journal entries (17.90 entries per participant on average), 0.62 entries on average per day (more than once every two days). 22 out of 28 participants used MindfulDiary more than once every two days. Participants generally engaged with the app at a regular frequency, but we note that their engagement was also affected by the three-day-miss reminder and their visit to the clinic between Week 2 and 4. Each journaling session lasted an average of 438 seconds (around 7 minutes) but with notable individual variability ($SD = 225.97$). Each journal dialogue included messages with an average of 105.6 syllable count ($SD = 49.41$). Our analysis did not reveal significant differences in either the participants' input length ($F(1.735, 46.85) = 2.718$, $p = .084$) or writing time ($F(2.417, 65.25) = 2.549$, $p = .076$) across the four different time points, as determined by the RM-ANOVA test. This suggests that users mostly retained a steady level of engagement during the four-week study.

6.2 Dialogue Patterns

Participants and MindfulDiary exchanged a total of 4,410 messages (*i.e.*, 2,205 pairs of the AI and participant's messages) during the field study. Each session consisted of 10.82 messages ($SD = 2.70$). Most exchanges between the AI and participants were carried on for an exploration of patients' daily lives and emotions, as well as for casual conversations. In terms of the stage of the conversation, 62% (2,732 messages) of the messages were from Exploration, 30% (1220 messages) for Rapport building, and 6% (282 messages) for Sensitive topic. Only a small amount of messages were accounted for Wrapping up (62 messages) or not selected (14 messages).

To understand the contents that MindfulDiary generated, we delved deep into the content it generated. 72% of the AI messages took the form of questions, aiming to elicit responses about users' daily experiences and emotions. We identified and categorized the primary strategies that MindfulDiary employed to assist patients' journaling. There were four strategies employed by the LLM: *Emotional Exploration*, *Activity/Behavior Exploration*, *In-depth Follow-up & Countermeasures*, and *Future Plan Exploration*. For a comprehensive breakdown of these strategies, along with their descriptions and exemplar questions, refer to Table 3.

The average length of participants' responses was 29.42 syllable counts, with a median of 20 ($SD = 35.9$). This suggests a left-skewed distribution, where many participants gave shorter responses and a smaller number provided considerably longer answers, causing a high variation. The minimum response length was one character, and the maximum was 559 syllable counts. We further conducted a qualitative analysis of these responses, seeking to identify the themes present in users' interactions with the LLM. This allowed us to understand the scope and topics of the daily records that MindfulDiary collected from the patients.

Participants interacting with MindfulDiary conveyed a range of topics (see Table 4). They described a spectrum of *emotional states*, from negative feelings like exhaustion and anxiety to positive sentiments of pride and joy. *Events and activities* were recounted, offering insights into their daily routines, such as walking during school times or decreased activity post-vacation. They also shared *thoughts and beliefs*, sometimes related to current events, revealing patterns linked to mental health, like feelings of exclusion and loneliness. Regarding *perceived health status*, comments spanned from immediate ailments, such as headaches, to long-term health challenges. Distorted perceptions about their body included content on excessive dieting. Specifically, participants frequently discussed medications, revealing not just their physical reactions but also their perceptions and behaviors toward them. Some expressed concerns over the taste, while others mentioned adverse reactions from intake, like discomfort after swallowing multiple pills at once. Lastly, the realm of *relationships & interactions* had participants highlighting both the challenges and supports in their interpersonal connections, revealing their significant impact on mental well-being, from conflicts and trust issues to moments of affirmation and encouragement.

6.3 Patients' Perspectives on MindfulDiary

Overall, participants viewed MindfulDiary as a space where they could open up and share their stories, feeling a sense of empathy from the system. Participants particularly found the dialogue-driven interactions with MindfulDiary useful. One participant, P15, mentioned, "*If it was just about recording daily activities or emotions like a regular diary, it might have been less engaging, and I could've found it tedious or might not have persisted for long. But this felt like having a conversation with AI, which added an element of fun and*

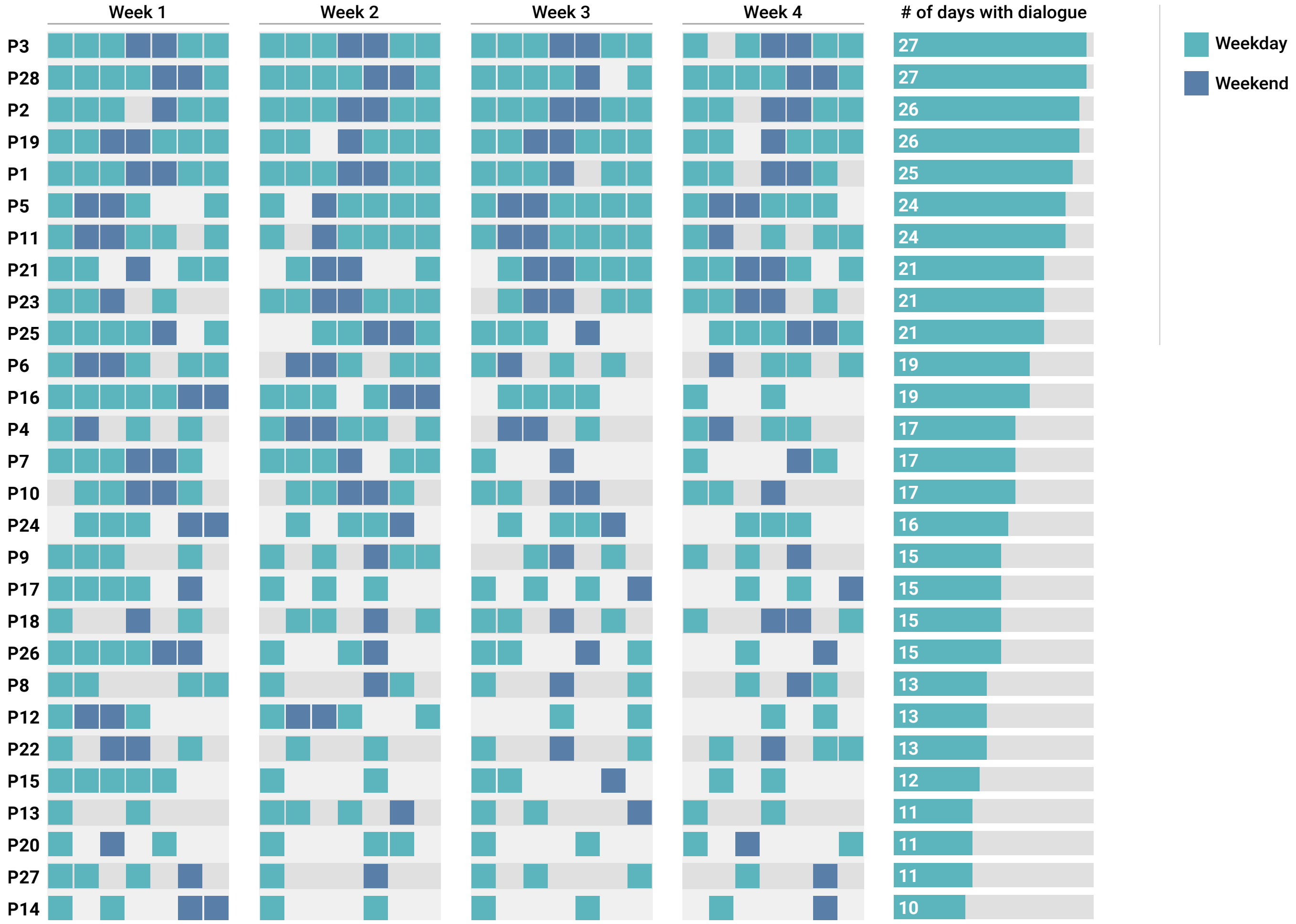


Figure 6: Overview of daily engagement of participants with MindfulDiary. The colored squares denote the days that participants conversed with MindfulDiary, with darker colors indicating weekend days. The bar charts on the right visualize the total number of days with interaction against the four-week study period. Participants are sorted by the number of days with interaction.

kept me engaged in continuous use." Such a dialogue-driven journaling process aided participants in maintaining consistent records and helped in forming a habit consistent with our user engagement analysis. P7 stated, "*I liked chatting with the AI at first, so I kept using it. The more I used it, the more it became a habit.*"

6.3.1 *Broad Conversational Range*: *The Versatility in Documenting Diverse Interests.*

Our participants appreciated the LLM's flexibility and naturalness in responding to various utterances, topics, and situations. Such broad conversational capabilities of the LLM provided participants with a space where they could document a variety of subjects tailored to individual interests and preferences. In our study, participants interacted with the LLM on diverse topics ranging from games, webcomics, novels, and movies (see Dialogue 1) to hobbies like Pilates (see Dialogue 2), allowing them to create richer and more personal records. P3 remarked, "*AI systems that I have used in the past could only respond to specific words, but it is amazing how this one can respond to all sorts of things.*"

P15	*In the afternoon, I watched movies with my family. We watched two films, Asura and The Witness, which were both quite interesting and thrilling.*
AI	*Watching movies with family probably made it more enjoyable. 'Asura' and 'The Witness', must have been exciting to watch, since they are both known for their strong stories!*

Dialogue 1. Dialogue snippet from P15

6.3.2 *Expanding Views*: *Enriching Entries with Varied Perspectives.*

Participants also valued the diverse and new perspectives that LLM-generated responses offered, as those helped participants reflect on their struggles, daily events, and emotions from various angles. Dialogue 3 shows how the system helps the participant to view the challenges of studying from the perspective of the satisfaction felt in gaming. This influence helped participants diverge from

Table 3: Categorization of LLM's strategies for generating prompts to encourage user reflection, including descriptions and exemplar questions. Strategies include Emotional Exploration, Activity Exploration, In-depth follow-up, and Future Plan Exploration with associated descriptions and examples

Category	Description	Example
Emotional exploration	Messages that further inquire about the user's emotional state, mood, and condition.	``How did you feel after meeting her?'' ``Can you tell me how this situation makes you feel?''
Activity exploration	Messages exploring the user's experiences or activities. These questions delve into more objective and factual content.	``What kind of exercise do you do?'' ``Lunch with a friend sounds nice. What did you eat?''
In-depth follow-up	Messages that help delve into the root causes or reasons related to mood, feeling, or specific situations.	``Since when have you felt this overshadowed mood?'' ``If you feel confused, what might be the reason?''
Future plan exploration	Messages inquiring about how the user reacted to or plans to respond to the events or emotions they mentioned.	``That situation must have been challenging. Have you considered any solutions?'' ``What have you done to alleviate the loneliness, even just a bit?''

Table 4: Summary of Participants' Input Messages: Categories include Emotional States, Events & Activities, Thoughts & Beliefs, Perceived Health Status, and Relationships & Interactions with associated descriptions and examples

Category	Description	Example
Emotional states	The emotions that participants documented in their daily lives encompassed a broad spectrum, ranging from negative and depressive sentiments.	"*I'm so exhausted, I feel like I'm reaching my limit soon.*" "*I'm very worried, scared, and anxious.*"
Events and activities	The mentions of events, tasks, or activities they participated in or witnessed, such as exam periods or travel.	"*When I attended school, I got some walking in, but after vacation, I don't have much reason to go out, leading to a decreased activity level.*"
Thoughts and beliefs	The thoughts, values, beliefs, and convictions they usually held. Including characteristic thought patterns related to mental health, such as distorted thinking.	"*I feel like someone is talking behind my back somewhere; they don't like my actions and seem to exclude me.*"
Perceived health status	The physical state and health concerns, misconceptions about their health, reactions to medications, revealing their perceptions, behaviors related to drug intake.	"*I'll starve and exercise to lose weight!*" "*I just took my medicine, but it seems to be getting tasteless.*"
Relationships & interactions	The contents about relationships with others. Difficulties stemming from interpersonal relationships, as well as the support and affirmation received from those around them.	"*I hated seeing my brother being happy. Forcing a cheerful tone also irked me.*"

P5	*I've been dancing to exercise, but I don't think it's enough.*
AI	*Dance is a fun and nice form of exercise. But if you feel it's lacking, Why do you think you feel that way?*
P5	*I want to exercise by learning from others, like in Pilates.*
AI	*You've become interested in Pilates. Can you tell me what makes Pilates interesting to you?*
P5	*I would like to do Pilates, but I can't because there's nowhere to do it.*
AI	*That must be a difficult situation. What about exercising at home? There are many Pilates videos available.*
P5	*I hadn't thought of following along with videos.*

Dialogue 2. Dialogue snippet from P5

ruminating on depressive feelings. P12 mentioned, "*Sometimes when you note down emotions, that's the only thought that comes to mind. Beyond that, I don't remember much. Since MindfulDiary uses AI, my thoughts flow more easily, and I like it when it asks about different perspectives or topics.*".

6.3.3 *Probing for Depth: Prompt Questions in Detailed Reflection.*

MindfulDiary's question-driven journaling process was also valued by participants as it assisted them with the process of daily reflections and documentation. Compared to their past experiences of journaling, where they had to reflect on their daily life by themselves, participants appreciated that MindfulDiary made the journaling process less daunting. P27 said, "*Because I have to rely solely on my thoughts when I write alone, I sometimes get stuck. But when I was unsure about how to write, the AI helped me. I liked that part.*". The questions posed by MindfulDiary also guided participants in documenting their daily lives in a more detailed manner by asking their thoughts and feelings about a particular event (See Dialogue 4).

AI	*Such ordinary days accumulate to form our daily lives. What emotions do you feelwhen you study or play games*?
P2	*I don't like studying, but I'm working hard dreaming about the college I want to go to. Games are, of course, fun and I feel proud when my character grows.*
AI	*So, studying can be seen as an investment for the future.*
P2	*Right, studying is an investment for the future.*
AI	*How does it feel when your game character grows*? *Can you bring that feeling to studying too*?
P2	*The feeling of my game character growing makes me proud, as if I've achieved something. I feel the same pride when I complete my daily study goals.*
AI	*How about continuing to study with that sense of pride*? *Could it bring the same sense of achievement*?
P2	*Based on this pride, I plan to focus more on my studies. It won't be easy, but setting clear goals can lead to faster achievements.*

Dialogue 3. Dialogue snippet from P2

Such probing allowed for richer, more in-depth entries. P13 shared, "*I used to write diaries on my own and just wrote whatever came to mind. MindfulDiary, however, helped me write in more detail because of the specific questions.*"

	[...]
P13	*I felt so pathetic and wanted to know why I was acting this way.*
AI	*It must have been really tough to go through such feelings. Does this happen often*?
P13	*It seems to have gotten worse recently.*
AI	*I see. Do you have any thoughts on why it's getting worse*?
P13	*Lately, I've been mentally exhausted and stressed a lot.*
AI	*That sounds really difficult. Can you tell me the sources of your stress*?
P13	*Issues with my dorm, relationships with my friends, and worries about the future.*
AI	*It must be hard with all these concerns overlapping. What's the most pressing issue for you right now*?
P13	*These days, I am very concerned about my relationships with friends.*
AI	*I understand your worries about friendships. What is your biggest concern*?*e*
P13	*I feel like people are talking behind my back, disliking my actions, and excluding me from gatherings, which makes me feel very lonely.*
	[...]

Dialogue 4. Dialogue snippet from P13

6.3.4 Building Narratives: Structuring Daily Reflections with MindfulDiary. MindfulDiary's capabilities, such as generating contextualized follow-up questions and summarizing the conversation, made the process easier for participants who struggled to organize daily thoughts and events underpsychotherapy [23]. In their past experiences, our participants expressed difficulties in journaling because of disjointed thoughts, a lack of clarity in ideas, or inconsistencies in their stories. However, with the support of the LLM in the MindfulDiary, these challenges were addressed, motivating them to record their daily lives persistently. P3 remarked, " *I often had trouble putting sentences In the past, I would worry about writing the next part. But with this tool, I just tell the story of my day, and it seamlessly continues and wraps it up, presenting a well-structured diary entry. That's its biggest advantage.*" (See Dialogue 5)

	[...]
P3	*The most memorable thing today was my dad eating the cake and food I made yesterday and saying it was delicious.*
AI	*How was the experience of making the cake*? *Did you enjoy it*?
P3	*The tough day turned to be enjoyable.*
AI	*Can you describe more about that feeling of happiness*?
P3	*I felt a bit disappointed and tired while doing it, but when my dad liked it, I felt like I did something great and it made me happy.*
	[...]
	[Generated Diary] ``Today was a day without any special worries. The most memorable thing was my dad eating and complimenting the cake and food I made yesterday. That moment turned a somewhat exhausting and difficult day into a joyful one. My father's praise gave me a great sense of joy and pride. It might be a wrong thought, but I felt really great, like I am an amazing person, which was so nice.''

Dialogue 5. Dialogue snippet from P3

6.4 MHPs' Perspectives on MindfulDiary for Clinical Mental Health Settings

In this section, we describe how MHPs utilized the clinician dashboard and the benefits and drawbacks of the system they reported, drawing on the debriefing interviews with the psychiatrists.

6.4.1 Utilization of MindfulDiary in Clinic. During the deployment study, psychiatrists reviewed the journal entries from their patients every morning when they reviewed the medical charts of patients whom they would meet on the day. Depending on the severity and the focal concerns of the patient, psychiatrists spent about 5 to 10 minutes per patient reviewing the MindfulDiary data. After checking trends primarily through PHQ-9 in the clinician dashboard, psychiatrists read summaries about events and documented emotions. If there were spikes or drops in the PHQ-9 or events/emotions, they checked the actual dialogues.

6.4.2 Percevied Benefits of MindfulDiary for Enhanced Patient Insight and Empathetic Engagement. All of the psychiatrists emphasized the critical value of an expert interface based on information recorded in the daily lives of patients. Specifically, E3 highlighted MindfulDiary's value in that it consistently aids in recording daily entries, allowing them to utilize more detailed patient data during outpatient visits. "*Patients, with the support of AI, can logically continue their narratives, ensuring more dialogue than a typical* (*paper-based*) *diary. This definitely aids me in my consultations.*" (E3). In this section, we further report on how MindfulDiary has been helpful in the clinical practice of psychiatrists.

Enhancing Understanding and Empathy toward Patients. Psychiatrists indicated that MindfulDiary helped them gain a deeper understanding and empathy about their patients. They perceived that MindfulDiary served as a questioner that could elicit more objective and genuine responses from patients. Psychiatrists appreciated that the LLM was able to pose questions that might be sensitive or burdensome for them to ask, such as patients' negative perceptions of their parents. E4 said: "*There are times when it's challenging to counter a patient's narrative or offer an opposing perspective. For example, if a patient speaks very negatively about their mother, and we ask, 'Didn't she treat you well when you were younger?', the patient might react aggressively, thinking, 'Why is the therapist taking my mother's side?' However, since the LLM is a machine, such concerns are minimized.*".

Insights from Everyday Perspectives Outside Clinical Visits. Psychiatrists valued that MindfulDiary provided them with an understanding of patients' conditions that would be difficult to gain during outpatient visits. For instance, E1 appreciated that MindfulDiary provided them with insights into patients' positive feelings and experiences, which is typically difficult to obtain during clinical consultations. "*Usually, when patients come for a consultation, they talk about bad experiences. Few people come to psychiatry to say, 'I've been doing well.' Even if they have good things to say, they usually don't bring them up. But I was happy to see that there were many positive statements in these records, like 'I did that and felt good.' Especially in depression, the presence or absence of positive emotions is crucial. It's a good sign if they show such positive responses.*". E2 envisioned its potential application to medication management, which is another critical aspect of psychiatric care. He thought these records could be used as a window into understanding how patients react to and perceive medications. For patients undergoing drug therapy, "*If the primary treatment method is pills, but they don't seem to have an effective response or there's a decline in medication acceptance, I could potentially understand the reasons for it through this diary.*" (E2).

Understanding Patient Progress Through Consistent Record-Keeping. Feedback from patients highlighted that interactions with MindfulDiary made it easier for patients to maintain a consistent record, as it mitigated the challenges associated with recording. Psychiatrists perceived that having consistent daily data offered them opportunities to observe trends in a patient's condition. E2 said: "*From our perspective as clinicians, even though we might only see a patient once a month, having access to a record of how they've been throughout the month would allow us to track their progress, which is highly beneficial.*". In particular, the ability to examine changes not only through quantitative tools like the PHQ-9 but also using a qualitative approach can offer a comprehensive understanding and shed light on the mechanisms influencing a patient's mental health.

6.4.3 Perceived Concerns about MindfulDiary.

While MHPs generally appraised the utility of the MindfulDiary positively, they also raised concerns regarding the integration of MindfulDiary into clinical settings.

Significance of Tone and Manner in Patient Data Analysis. Although patient data summarized and extracted in the expert interface effectively aided in understanding the patient, experts thought that the summarized texts would not convey the patient's tone, pace, and other nuances, which are integral to the Mental Status Examination (MSE) that clinicians utilize. However, MHPs identified the opportunity to perform such analysis from the raw data that patients entered. As the MSE measures objective and quantitative aspects, incorporating such an analysis could make significant improvements in understanding the patient. E1 said, "*In the same way as P14, understanding the tone of this patient may also be possible. That's because we use something called psychiatric MSE, where we observe more than just the patient's appearance, such as tone, pace, and more. Even a short analysis of one's linguistic behavior would be great.*"

Potential Misuses and Concerns around MindfulDiary. In our field study, one patient participant perceived the MindfulDiary as a channel to convey their intentions and situations to their psychiatrist. Specifically, the participant, P9, talked to their psychiatrist, "*Have you seen what I wrote?*", which indicated that the patient was actively attempting to share their current state and situation through MindfulDiary. In spite of the fact that such usage did not seem problematic per se, one psychiatrist raised concerns about the possibility that patients with borderline personality disorders might misuse MindfulDiary as a weapon to manipulate others, such as their providers and parents. "*In some cases, people self-harm out of genuine distress, but others do it to manipulate others, instilling guilt in them so they'll do what they want. There are some patients who write about their distress with sincerity, while there are some who exaggerate their distress in order to get attention.*" For patients exhibiting symptoms of schizophrenia or delusions, there was a concern that MindfulDiary's feature of revisiting past entries could act as a feedback loop, developing and amplifying their delusions. E2 said, "*This diary lets you revisit and organize your past actions. For schizophrenia patients with delusions or unique beliefs, referencing past writings might reinforce their pre-existing delusions. Reaffirming 'Yes, I'm right' can be problematic. The LLM's summaries could exacerbate these delusions if they emphasize distorted content.*"

7 DISCUSSION

In this study, we present MindfulDiary, an LLM-driven journal designed to document the daily experiences of psychiatric patients through naturalistic conversations. Here, we reflect on the opportunities presented by LLM-driven journaling for psychiatric patients and discuss considerations for integrating an LLM-driven patient system into the clinical setting.

7.1 Guiding Patient Journaling through Conversations Offering Diverse Perspectives

Our study highlighted the potential of MindfulDiary in clinical settings, mainly where adherence to interventions is important [62]. Core symptoms of depression, such as loss of energy, difficulty in carrying out mental processes, and feelings of apathy, often contribute to lower adherence to a professional' s advice or intervention [43]. Clinicians who participated in our FGI also highlighted these challenges in motivating patients to utilize the diary writing app. Our findings demonstrated that MindfulDiary helped mitigate these challenges by transforming the conventional journaling process into engaging conversations. Using MindfulDiary, users were able to engage in conversations with the system by answering prompts and questions, which made them feel the journaling process was more accessible and intriguing. This active participation ensures that the users are not overwhelmed by the task and are guided in documenting their feelings and experiences more richly.

Depression often locks patients into negative and rigid thought patterns [12]. Such patterns, resistance to change established thought paradigms, can severely limit a patient's ability to perceive issues from multiple angles, leading to a harsh self-judgment [61]. Our study highlighted that the varied perspectives offered by LLM-driven chatbots like MindfulDiary could help challenge such fixed viewpoints [33]. By prompting users to revisit their initial evaluations or suggest alternative viewpoints, these chatbots could help break the cycle of cognitive rigidity. While our research underscores the promising role of LLM-driven chatbots in assisting psychiatric patients' journaling process, it's essential to note that these are preliminary findings. More work is needed to substantiate these findings in a clinical context.

7.2 MindfulDiary as a Facilitator for Fostering Patient-Provider Communication

Studies have suggested that sharing the data captured via chatbots with others, such as health professionals and family members, could further serve as an effective mediator that helps convey more truthful information [52, 56]. For instance, patients consistently displayed deep self-disclosure through chatbots, whether or not they intended to share their inputs with health professionals [52]. Aligned with prior work on PGHD [20, 64], MHPs in our study also perceived that MindfulDiary has shed light on patients' daily events, emotions, and thoughts that might have been difficult to gain through regular clinical visits. This data offered MHPs valuable insights into the patient's experiences and context.

Building on these findings, we could expand the potential presented by MindfulDiary in patient-provider communication. In the field of personal health informatics, existing research highlights the role of technology, such as photo journaling, in managing conditions like Irritable Bowel Syndrome. This tool not only empowers patients to record their daily experiences more effectively but also fosters enhanced collaboration between patients and healthcare providers [18, 79]. Such tools serve as vital artifacts in negotiating the boundaries of patient-provider interactions (i.e., boundary negotiating artifacts) [19].

This work adds a new dimension to this discussion by showing how LLM-assisted journaling lowers barriers to generating health data in daily life and fosters patient understanding. Specifically, we found that through this system, patients and providers can collaboratively reflect on mental health conditions. In the context of the stage-based model of personal informatics, the patient module in our MindfulDiary helps patients reduce the burden of collecting daily data and supports deeper recording. The expert module's dashboard allows for the combined and transformed processes of diary data, survey data, and quantitative engagement data, supporting MHPs' integration and reflection. Collaborative data generation and utilization with patients can enable care that reflects the patient's values and the characteristics of their daily life. These insights serve as a basis for patient-provider collaboration.

However, our study findings underscore the importance of careful consideration in the clinical integration of systems like MindfulDiary. While we did not observe patients exaggerating their conditions or needs, this potential issue was raised as a concern by MHPs. They expressed apprehension about the possibility that sharing journal content with MHPs through MindfulDiary might lead some patients to exaggerate their conditions or needs. This concern highlights the need to consider not only the design of chatbots that facilitate patient disclosure behavior but also the complex dynamics between patients and providers in clinical settings. It is crucial to address these dynamics to ensure the effective and safe use of such technologies in mental health care. The growing prevalence of chatbots in mental health domains emphasizes the need for a holistic approach to their design and implementation. We highlight that engineers and MHPs need to collaborate closely, ensuring that these tools are not only technically sound but also tailored to meet the intricate dynamics of clinical settings [87].

7.3 Considerations for Integrating LLMs into Clinical Settings

In this section, we discuss the consideration for integrating LLMs into clinical mental health settings, drawing insights from the design and evaluation of MindfulDiary.

Aligning Domain Experts' Expectations of LLMs. Developing and deploying MindfulDiary, we learned that aligning MHPs' expectations with the capabilities and limitations of LLMs involves significant challenges. The capability of generative language models to improve mental health is difficult to measure in comparison with AI models in other medical domains, where objective metrics can determine performance. For instance, in medical imaging, AI can be evaluated based on its accuracy in identifying target diseases from MRI scans, using precise numerical percentages of correct identifications [83]. On the other hand, in the realm of mental health chatbots, gauging success is more nuanced, as it involves subjective interpretations of emotional well-being and psychological improvement, which cannot be easily quantified or compared in the same straightforward manner. This challenge is amplified in mental health, where soft skills like rapport building and emotional observation are important [30]. The use of LLMs in the mental health field is emerging, but little has been said about evaluating or defining the performance of models that are tailored to mental health. Our iterative evaluation process involving MHPs could inform researchers about how to develop and evaluate LLM-mediated mental health technology. When integrating into the clinical setting, this evaluation is also necessary for anticipating who the system would target and for what purpose it would be used. Hence, we advocate that engineers and researchers should carefully consider how to assist domain experts, who may lack AI expertise, in fully and accurately grasping the role and operation of

LLM. It is also crucial for researchers and engineers to collaborate closely with these professionals to ensure the technology aligns with therapeutic needs and best practices [87].

Tailored LLM Evaluation for Clinical Mental Health Domains. The domain of mental health, which our study addresses, is characterized by the vulnerability of its target user group. The content discussed within this domain is often emotionally charged and sensitive. Therefore, prioritizing user safety becomes even more essential in this domain than in others. Considering the sensitivity of the domain, during our evaluation process, MHPs thoroughly tested the LLM's output by trying out conversations on various sensitive topics in both implicit and explicit ways, drawing upon their clinical experiences. The contents the MHPs input were much more diverse and wide-ranging than what engineers could generate during the development. Additionally, MHPs showed concern that the hallucinations of the LLM could reinforce or expand the delusions of patients with delusional disorders. We highlight that developing evidence-based tests or benchmark sets to anticipate the behavior of the language models in collaboration with MHPs is critical when leveraging LLMs for clinical mental health settings.

Incorporating Perspectives of MHPs in Testing and Monitoring. Considering the caveats of current LLMs [47], it is critical to involve MHPs when deploying LLM-driven systems for patients in mental health contexts. While planning the field deployment study of MindfulDiary, we identified specific roles that MHPs could play. In the pre-use phase, MHPs should determine the suitability of users and facilitate the onboarding process with patients. During the mid-use phase, they should closely monitor interactions with the LLM and be prepared to intervene in cases of crises or unexpected use scenarios. Furthermore, they can offer or adjust treatments periodically based on long-term data. Additionally, they should regularly re-evaluate the continued use of the system. While some of these tasks should carefully be designed not to burden MHPs too much, it is important that LLMs do not make autonomous decisions about patients (e.g., diagnosis, prescription, or crisis management) but instead operate under professional oversight.

Providing Safeguards for Hallucinated LLM Generations. Our clinician dashboard provided various summarized information, such as word cloud, aggregating multiple dialogue entries so that the clinician quickly grasps the gist of the dialogues. Although we underwent intensive testing with the LLM-driven data summarizer, the LLM-driven data processing may still suffer from inaccuracies, biases, and misinterpretation [40, 75] of patient sentiments or context, which could adversely affect treatment decisions and patient well-being. To mitigate such drawbacks of LLMs in our study, we provided sufficient guidance to MHPs, cautioning them that the LLM-generated information they receive may be error-prone. However, in real-world settings, MHPs might accept the outputs of LLM without much attention. Therefore, when involving LLM-driven data processing, the system should foster careful reviewing of the content based on the expertise of MHPs. For example, future systems could incorporate features like highlighting *in vivo* phrases that were directly mentioned by patients and signify key aspects of their experience and feelings. By contrasting the *in vivo* phrases with the LLM's original text, the system can encourage MHPs to put more scrutiny on the LLM's original interpretation, which may contain errors, and the actual inputs spoken by patients.

7.4 Limitations and Future Work

Our recruitment method could impact the generalizability of our findings, as we recruited the patient participants for our field study from a single university hospital. Although we aimed to recruit patients with diverse types and levels of symptoms, our participants are not representative samples of psychiatric patients. They were young (mostly adolescents) and consulted by a fixed number of psychiatrists. While this work is just a first step toward designing an LLM-driven journaling app for psychiatric patients, further investigation is necessary with subjects from various backgrounds. To implement our pipeline, we used OpenAI's GPT API, which provided the most capable LLM at the time of our study and was accessible via commercial API. As GPT models are continually updated, later models may not yield the same conversational behavior. To generalize the performance of our conversational pipeline design, future work is needed to compare multiple versions of MindfulDiary with different underlying LLMs.

8 CONCLUSION

In this paper, we designed MindfulDiary to assist psychiatric patients undergoing outpatient treatment with journaling in their daily lives. Keeping the clinical mental health setting in mind, our system was developed in collaboration with MHPs, from the initial concept building to the design of LLM's conversation flow and evaluation. MindfulDiary leverages a stage-based LLM-driven chatbot, enabling patients to interact through prompt questions and answers, while complying with guidelines based on MHPs and literature. We conducted a field deployment study with 28 patients over 4 weeks. We found that the versatility, narrative-building capability, and diverse perspectives provided by MindfulDiary assisted patients in consistently enriching their daily records. The enriched records from MindfulDiary provided psychiatrists with deeper insights, enhancing their understanding and empathy toward their patients. We hope that this research provides a case study and insight into the development of an LLM-driven chatbot for mental health that is clinically relevant and reflects the needs and experiences of MHPs.

ACKNOWLEDGMENTS

We thank our study participants for their time and efforts. We also thank Eunkyung Jo and Yubin Choi for providing feedback on the early draft of this paper. This work was supported through a research internship at NAVER AI Lab, and in part supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI22C1962) and by the National Research Foundation of Korea grant funded by the Ministry of Science and ICT (No. 2022R1G1A1009100).

REFERENCES

[1] Alaa A. Abd-alrazaq, Mohannad Alajlani, Ali Abdallah Alalwan, Bridgette M. Bewick, Peter Gardner, and Mowafa Househ. 2019. An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics* 132 (2019), 103978. https://doi.org/10.1016/j.ijmedinf.2019.103978

[2] Alaa A Abd-Alrazaq, Mohannad Alajlani, Ali Abdallah Alalwan, Bridgette M Bewick, Peter Gardner, and Mowafa Househ. 2019. An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics* 132 (2019), 103978.

[3] Alaa A Abd-Alrazaq, Mohannad Alajlani, Nashva Ali, Kerstin Denecke, Bridgette M Bewick, and Mowafa Househ. 2021. Perceptions and opinions of patients about mental health chatbots: scoping review. *Journal of medical Internet research* 23, 1 (2021), e17828.

[4] Apple Inc. 2024. Getting Started with the Weather App on iPhone. https://support.apple.com/ko-kr/guide/iphone/iph0e5ca7dd3/ios. Accessed: 2024-02-15.

[5] Riku Arakawa, Hiromu Yakura, and Masataka Goto. 2023. CatAlyst: Domain-Extensible Intervention for Preventing Task Procrastination Using Large Generative Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany </country>, </conf-loc>) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 157, 19 pages. https://doi.org/10.1145/3544548.3581133

[6] Elizabeth Austin, Jenney R Lee, Dagmar Amtmann, Rich Bloch, Sarah O Lawrence, Debbe McCall, Sean Munson, and Danielle C Lavallee. 2020. Use of patient-generated health data across healthcare settings: implications for health systems. *JAMIA open* 3, 1 (2020), 70–76.

[7] John W Ayers, Adam Poliak, Mark Dredze, Eric C Leas, Zechariah Zhu, Jessica B Kelley, Dennis J Faix, Aaron M Goodman, Christopher A Longhurst, Michael Hogarth, et al. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine* (2023).

[8] Sanghwan Bae, Donghyun Kwak, Sungdong Kim, Donghoon Ham, Soyoung Kang, Sang-Woo Lee, and Woomyoung Park. 2022. Building a role specified open-domain dialogue system leveraging large-scale language models. *arXiv preprint arXiv:2205.00176* (2022).

[9] Karen A. Baikie, Liesbeth Geerligs, and Kay Wilhelm. 2012. Expressive writing and positive writing for participants with mood disorders: An online randomized controlled trial. *Journal of Affective Disorders* 136, 3 (2012), 310–319. https://doi.org/10.1016/j.jad.2011.11.032

[10] David Bakker and Nikki Rickard. 2018. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. *Journal of affective disorders* 227 (2018), 432–442.

[11] Andrew L Beam and Isaac S Kohane. 2016. Translating artificial intelligence into clinical care. *Jama* 316, 22 (2016), 2368–2369.

[12] Aaron T Beck. 1979. *Cognitive therapy and the emotional disorders*. Penguin.

[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[14] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[15] Rafael A Calvo, David N Milne, M Sazzad Hussain, and Helen Christensen. 2017. Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering* 23, 5 (2017), 649–685.

[16] Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. 2019. Assessing the usability of a chatbot for mental health care. In *Internet Science: INSCI 2018 International Workshops, St. Petersburg, Russia, October 24–26, 2018, Revised Selected Papers 5*. Springer, 121–132.

[17] Seong-Jin CHO, Hong-Jin JEON, Moo-Jin KIM, Jang-Kyu KIM, Uk-Seon KIM, In-Kyoon LYOO, and Maeng-Je CHO. 2001. Prevalence and correlates of depressive symptoms among the adolescents in an urban area in Korea. *Journal of Korean Neuropsychiatric Association* (2001), 627–639.

[18] Chia-Fang Chung. 2017. Supporting Patient-Provider Communication and Engagement with Personal Informatics Data. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) (*UbiComp '17*). Association for Computing Machinery, New York, NY, USA, 335–338. https://doi.org/10.1145/3123024.3123197

[19] Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine Zia, James Fogarty, Julie A. Kientz, and Sean A. Munson. 2016. Boundary Negotiating Artifacts in Personal Informatics: Patient-Provider Collaboration with Patient-Generated Data. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) (*CSCW '16*). Association for Computing Machinery, New York, NY, USA, 770–786. https://doi.org/10.1145/2818048.2819926

[20] Deborah J Cohen, Sara R Keller, Gillian R Hayes, David A Dorr, Joan S Ash, and Dean F Sittig. 2016. Integrating patient-generated health data into clinical care settings or clinical decision-making: lessons learned from project healthdesign. *JMIR human factors* 3, 2 (2016), e5919.

[21] George Demiris, Sarah J Iribarren, Katherine Sward, Solim Lee, and Rumei Yang. 2019. Patient generated health data use in clinical practice: a systematic review. *Nursing Outlook* 67, 4 (2019), 311–330.

[22] Dina Demner-Fushman, Wendy W Chapman, and Clement J McDonald. 2009. What can natural language processing do for clinical decision support? *Journal of biomedical informatics* 42, 5 (2009), 760–772.

[23] Daniel A Donnelly and Edward J Murray. 1991. Cognitive and emotional changes in written essays and therapy interviews. *Journal of Social and Clinical psychology* 10, 3 (1991), 334–350.

[24] Chris Elsden, Abigail C. Durrant, and David S. Kirk. 2016. It's Just My History Isn't It? Understanding Smart Journaling Practices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 2819–2831. https://doi.org/10.1145/2858036.2858103

[25] Zohar Elyoseph, Dorit Hadar-Shoval, Kfir Asraf, and Maya Lvovsky. 2023. ChatGPT outperforms humans in emotional awareness evaluations. *Frontiers in Psychology* 14 (2023), 1199058.

[26] Faiza Farhat. 2023. ChatGPT as a complementary mental health resource: a boon or a bane. *Annals of Biomedical Engineering* (2023), 1–4.

[27] Mayara Costa Figueiredo, Yunan Chen, et al. 2020. Patient-generated health data: dimensions, challenges, and open questions. *Foundations and Trends® in Human–Computer Interaction* 13, 3 (2020), 165–297.

[28] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health* 4, 2 (2017), e19. https://doi.org/10.2196/mental.7785

[29] Russell Fulmer, Angela Joerin, Breanna Gentile, Lysanne Lakerink, Michiel Rauws, et al. 2018. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR mental health* 5, 4 (2018), e9782.

[30] Glen O. Gabbard and Holly Crisp-Han. 2017. The Early Career Psychiatrist and the Psychotherapeutic Identity. *Academic Psychiatry* 41, 1 (2017), 30–34. https://doi.org/10.1007/s40596-016-0627-7

[31] Jianfeng Gao, Michel Galley, and Lihong Li. 2018. Neural Approaches to Conversational AI. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*. Association for Computational Linguistics, Melbourne, Australia, 2–7. https://doi.org/10.18653/v1/P18-5002

[32] Simret Araya Gebreegziabher, Zheng Zhang, Xiaohang Tang, Yihao Meng, Elena L. Glassman, and Toby Jia-Jun Li. 2023. PaTAT: Human-AI Collaborative Qualitative Coding with Explainable Interactive Rule Synthesis. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3544548.3581352

[33] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing using Language Models. *Designing Interactive Systems Conference* (2022), 1002–1019. https://doi.org/10.1145/3532106.3533533

[34] Amy L. Gonzales, Tiffany Y. Ng, OJ Zhao, and Geri Gay. 2010. Motivating Expressive Writing with a Text-to-Sound Application. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) (*CHI '10*). Association for Computing Machinery, New York, NY, USA, 1937–1940. https://doi.org/10.1145/1753326.1753618

[35] Adam J. Guastella and Mark R. Dadds. 2009. Sequential Growth in Cognitive-behavioral Emotion-processing: A Laboratory Study. *Cognitive Therapy and Research* 33, 4 (2009), 368–374. https://doi.org/10.1007/s10608-008-9199-5

[36] Kent D Harber, James W Pennebaker, and S Christianson. 1992. Overcoming traumatic memories. *The handbook of emotion and memory: Research and theory* (1992), 359–387.

[37] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: A retrospective memory aid. In *UbiComp 2006: Ubiquitous Computing: 8th International Conference, UbiComp 2006 Orange County, CA, USA, September 17-21, 2006 Proceedings 8*. Springer, 177–193.

[38] Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in Building Intelligent Open-Domain Dialog Systems. *ACM Trans. Inf. Syst.* 38, 3, Article 21 (apr 2020), 32 pages. https://doi.org/10.1145/3383123

[39] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. 2018. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth* 6, 11 (2018), e12106. https://doi.org/10.2196/12106

[40] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *Comput. Surveys* 55, 12 (2023), 1–38.

[41] Eunkyung Jo, Daniel A. Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 18, 16 pages. https://doi.org/10.1145/3544548.3581503

[42] Vaiva Kalnikaite and Steve Whittaker. 2012. Synergetic recollection: how to design lifelogging tools that help locate the right information. In *Human-computer interaction: the agency perspective*. Springer, 329–348.

[43] Sidney H Kennedy. 2022. Core symptoms of major depressive disorder: relevance to diagnosis and treatment. *Dialogues in clinical neuroscience* (2022).

[44] Junhan Kim, Yoojung Kim, Byungjoon Kim, Sukyung Yun, Minjoon Kim, and Joongseek Lee. 2018. Can a Machine Tend to Teenagers' Emotional Needs? A Study with Conversational Agents. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI EA '18*). Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188548

[45] Kitty Klein and Adriel Boals. 2001. Expressive Writing Can Increase Working Memory Capacity. *Journal of Experimental Psychology: General* 130, 3 (2001), 520–533. https://doi.org/10.1037/0096-3445.130.3.520

[46] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 70 (jul 2018), 26 pages. https://doi.org/10.1145/3214273

[47] Diane M. Korngiebel and Sean D. Mooney. 2021. Considering the possibilities and pitfalls of Generative Pre-trained Transformer 3 (GPT-3) in healthcare delivery. *npj Digital Medicine* 4, 1 (2021), 93. https://doi.org/10.1038/s41746-021-00464-x

[48] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2001. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine* 16, 9 (2001), 606–613.

[49] Richard D Lane, Donald M Quinlan, Gary E Schwartz, Pamela A Walker, and Sharon B Zeitlin. 1990. The Levels of Emotional Awareness Scale: A cognitive-developmental measure of emotion. *Journal of personality assessment* 55, 1-2 (1990), 124–134.

[50] Minchul Lee. 2022. *Kiwi, Korean Intelligent Word Identifier.* https://doi.org/10.5281/zenodo.7041425

[51] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300932

[52] Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 31 (may 2020), 27 pages. https://doi.org/10.1145/3392836

[53] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-Disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376175

[54] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *Comput. Surveys* 55, 9 (2023), 1–35.

[55] Ross J Lordon, Sean P Mikles, Laura Kneale, Heather L Evans, Sean A Munson, Uba Backonja, and William B Lober. 2020. How patient-generated health data and patient-reported outcomes affect patient–clinician relationships: A systematic review. *Health Informatics Journal* 26, 4 (2020), 2689–2706. https://doi.org/10.1177/1460458220928184

[56] Kai Lukoff, Taoxi Li, Yuan Zhuang, and Brian Y Lim. 2018. TableChat: mobile food journaling to facilitate family support for healthy eating. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–28.

[57] Olivier Luminet, Bernard Rimé, R Michael Bagby, and Graeme Taylor. 2004. A multimodal investigation of emotional responding in alexithymia. *Cognition and emotion* 18, 6 (2004), 741–766.

[58] Haiwei Ma, C. Estelle Smith, Lu He, Saumik Narayanan, Robert A. Giaquinto, Roni Evans, Linda Hanson, and Svetlana Yarosh. 2017. Write for Life: Persisting in Online Health Communities through Expressive Writing and Social Support. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 73 (dec 2017), 24 pages. https://doi.org/10.1145/3134708

[59] Zilin Ma, Yiyang Mei, and Zhaoyuan Su. 2023. Understanding the Benefits and Challenges of Using Large Language Model-based Conversational Agents for Mental Well-being Support. *arXiv* (2023). https://doi.org/10.48550/arxiv.2307.15810 arXiv:2307.15810

[60] Farah Magrabi, Elske Ammenwerth, Jytte Brender McNair, Nicolet F De Keizer, Hannele Hyppönen, Pirkko Nykänen, Michael Rigby, Philip J Scott, Tuulikki Vehko, Zoie Shui-Yee Wong, et al. 2019. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. *Yearbook of medical informatics* 28, 01 (2019), 128–134.

[61] Donatella Marazziti, Giorgio Consoli, Michela Picchetti, Marina Carlini, and Luca Faravelli. 2010. Cognitive impairment in major depression. *European journal of pharmacology* 626, 1 (2010), 83–86.

[62] Peter Musiat, Catherine Johnson, Melissa Atkinson, Simon Wilksch, and Tracey Wade. 2022. Impact of guidance on intervention adherence in computerised interventions for mental health problems: a meta-analysis. *Psychological Medicine* 52, 2 (2022), 229–240. https://doi.org/10.1017/S0033291721004621

[63] Jodie Nghiem, Daniel A Adler, Deborah Estrin, Cecilia Livesey, Tanzeem Choudhury, et al. 2023. Understanding Mental Health Clinicians' Perceptions and Concerns Regarding Using Passive Patient-Generated Health Data for Clinical Decision-Making: Qualitative Semistructured Interview Study. *JMIR Formative Research* 7, 1 (2023), e47380.

[64] Shantanu Nundy, Chen-Yuan E Lu, Patrick Hogan, Anjuli Mishra, and Monica E Peek. 2014. Using patient-generated health data from mobile technologies for diabetes self-management support: provider perspectives from an academic medical center. *Journal of diabetes science and technology* 8, 1 (2014), 74–82.

[65] OpenAI, Inc. 2022. ChatGPT. https://chat.openai.com

[66] Ekkehard Othmer and Sieglinde C Othmer. 2002. *The clinical interview using DSM-IV-TR: Vol 1: Fundamentals.* American Psychiatric Publishing, Inc.

[67] Ekkehard Othmer and Sieglinde C Othmer. 2002. *The clinical interview using DSM-IV-TR: Vol. 2: The difficult patient.* American Psychiatric Publishing, Inc.

[68] SoHyun Park, Anja Thieme, Jeongyun Han, Sungwoo Lee, Wonjong Rhee, and Bongwon Suh. 2021. "I wrote as if I were telling a story to someone I knew.": Designing Chatbot Interactions for Expressive Writing in Mental Health. *Designing Interactive Systems Conference 2021* (2021), 926–941. https://doi.org/10.1145/3461778.3462143

[69] James W Pennebaker. 1985. Traumatic experience and psychosomatic disease: Exploring the roles of behavioural inhibition, obsession, and confiding. *Canadian Psychology/Psychologie canadienne* 26, 2 (1985), 82.

[70] James W Pennebaker. 1997. Writing about emotional experiences as a therapeutic process. *Psychological science* 8, 3 (1997), 162–166.

[71] James W Pennebaker and Cindy K Chung. 2011. Expressive writing: Connections to physical and mental health. (2011).

[72] James W Pennebaker and Janel D Seagal. 1999. Forming a story: The health benefits of narrative. *Journal of clinical psychology* 55, 10 (1999), 1243–1254.

[73] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[74] Gerardo Ramirez and Sian L Beilock. 2011. Writing about testing worries boosts exam performance in the classroom. *science* 331, 6014 (2011), 211–213.

[75] Vipula Rawte, Amit Sheth, and Amitava Das. 2023. A survey of hallucination in large foundation models. *arXiv preprint arXiv:2309.05922* (2023).

[76] Tara Riddell, Jane Nassif, Ana Hategan, and Joanna Jarecki. 2020. Healthy Habits: Positive Psychology, Journaling, Meditation, and Nature Therapy. *Humanism and Resilience in Residency Training: A Guide to Physician Wellness* (2020), 439–472.

[77] Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. Recipes for Building an Open-Domain Chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Association for Computational Linguistics, Online, 300–325. https://doi.org/10.18653/v1/2021.eacl-main.24

[78] Peter Salovey and John D Mayer. 1990. Emotional intelligence. *Imagination, cognition and personality* 9, 3 (1990), 185–211.

[79] Jessica Schroeder, Jane Hoffswell, Chia-Fang Chung, James Fogarty, Sean Munson, and Jasmine Zia. 2017. Supporting Patient-Provider Collaboration to Identify Individual Triggers Using Food and Symptom Journals. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) (*CSCW '17*). Association for Computing Machinery, New York, NY, USA, 1726–1739. https://doi.org/10.1145/2998181.2998276

[80] Woosuk Seo, Chanmo Yang, and Young-Ho Kim. 2023. ChaCha: Leveraging Large Language Models to Prompt Children to Share Their Emotions about Personal Events. arXiv:2309.12244 [cs.HC]

[81] Michael Shapiro, Douglas Johnston, Jonathan Wald, and Donald Mon. 2012. Patient-generated health data. *RTI International, April* 813 (2012), 814.

[82] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. 2023. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence* 5, 1 (2023), 46–57.

[83] Dinggang Shen, Guorong Wu, and Heung-Il Suk. 2017. Deep learning in medical image analysis. *Annual review of biomedical engineering* 19 (2017), 221–248.

[84] Vera Sorin, Danna Brin, Yiftach Barash, Eli Konen, Alexander Charney, Girish Nadkarni, and Eyal Klang. 2023. Large language models (llms) and empathy-a systematic review. *medRxiv* (2023), 2023–08.

[85] Robert L Spitzer, Kurt Kroenke, Janet BW Williams, and Bernd Löwe. 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine* 166, 10 (2006), 1092–1097.

[86] Michael J. Tanana, Christina S. Soma, Patty B. Kuo, Nicolas M. Bertagnolli, Aaron Dembe, Brian T. Pace, Vivek Srikumar, David C. Atkins, and Zac E. Imel. 2021. How do you feel? Using natural language processing to automatically rate

emotion in psychotherapy. *Behavior Research Methods* 53, 5 (2021), 2069–2082. https://doi.org/10.3758/s13428-020-01531-z

[87] Anja Thieme, Maryann Hanratty, Maria Lyons, Jorge E Palacios, Rita Marques, Cecily Morrison, and Gavin Doherty. 2022. Designing Human-Centered AI for Mental Health: Developing Clinically Relevant Applications for Online CBT Treatment. *ACM Transactions on Computer-Human Interaction* (2022). https://doi.org/10.1145/3564752

[88] John Torous, Nicole M Benson, Keris Myrick, and Gunther Eysenbach. 2023. Focusing on Digital Research Priorities for Advancing the Access and Quality of Mental Health. *JMIR Mental Health* 10 (April 2023), e47898. https://doi.org/10.2196/47898

[89] Cheryl Travers. 2011. Unveiling a reflective diary methodology for exploring the lived experiences of stress and coping. *Journal of Vocational Behavior* 79, 1 (2011), 204–216. https://doi.org/10.1016/j.jvb.2010.11.007

[90] Philip M. Ullrich and Susan K. Lutgendorf. 2002. Journaling about stressful events: Effects of cognitive processing and emotional expression. *Annals of Behavioral Medicine* 24, 3 (2002), 244–250. https://doi.org/10.1207/s15324796abm2403_10

[91] Allison Utley and Yvonne Garza. 2011. The Therapeutic Use of Journaling With Adolescents. *Journal of Creativity in Mental Health* 6, 1 (2011), 29–41. https://doi.org/10.1080/15401383.2011.557312

[92] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[93] Lu Wang, Munif Ishad Mujib, Jake Williams, George Demiris, and Jina Huh-Yoo. 2021. An Evaluation of Generative Pre-Training Model-based Therapy Chatbot for Caregivers. *arXiv* (2021). https://doi.org/10.48550/arxiv.2107.13115 arXiv:2107.13115

[94] Jing Wei, Sungdong Kim, Hyunhoon Jung, and Young-Ho Kim. 2024. Leveraging Large Language Models to Power Chatbots for Collecting User Self-Reported Data. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 87 (apr 2024), 35 pages. https://doi.org/10.1145/3637364

[95] Danny TY Wu, Chen Xin, Shwetha Bindhu, Catherine Xu, Jyoti Sachdeva, Jennifer L Brown, and Heekyoung Jung. 2020. Clinician perspectives and design implications in using patient-generated health data to improve mental health practices: mixed methods study. *JMIR formative research* 4, 8 (2020), e18123.

[96] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (*CHI '22*). Association for Computing Machinery, New York, NY, USA, Article 385, 22 pages. https://doi.org/10.1145/3491102.3517582

[97] J.D. Zamfirescu-Pereira, Heather Wei, Amy Xiao, Kitty Gu, Grace Jung, Matthew G Lee, Bjoern Hartmann, and Qian Yang. 2023. Herding AI Cats: Lessons from Designing a Chatbot by Prompting GPT-3. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (, Pittsburgh, PA, USA,) (*DIS '23*). Association for Computing Machinery, New York, NY, USA, 2206–2220. https://doi.org/10.1145/3563657.3596138

[98] Jian Zhao, Elaine Chapman, Stephen Houghton, and David Lawrence. 2022. Development and validation of a Coping Strategies Scale for use in Chinese contexts. *Frontiers in Psychology* 13 (2022), 845769.

[99] Zoie Zhao, Sophie Song, Bridget Duah, Jamie Macbeth, Scott Carter, Monica P Van, Nayeli Suseth Bravo, Matthew Klenk, Kate Sick, and Alexandre L. S. Filipowicz. 2023. More Human than Human: LLM-Generated Narratives Outperform Human-LLM Interleaved Narratives. In *Proceedings of the 15th Conference on Creativity and Cognition* (Virtual Event, USA) (*C&C '23*). Association for Computing Machinery, New York, NY, USA, 368–370. https://doi.org/10.1145/3591196.3596612

[100] Haining Zhu, Joanna Colgan, Madhu Reddy, and Eun Kyoung Choe. 2016. Sharing patient-generated data in clinical practices: an interview study. In *AMIA Annual Symposium Proceedings*, Vol. 2016. American Medical Informatics Association, 1303.