Missed Opportunities for Human-Centered AI Research: Understanding Stakeholder Collaboration in Mental Health AI Research

DONG WHI YOO, Kent State University, USA
HAYOUNG WOO, Georgia Institute of Technology, USA
SACHIN R. PENDSE, Georgia Institute of Technology, USA
NATHANIEL YOUNG LU, Zucker Hillside Hospital, Psychiatry Research, USA
MICHAEL L. BIRNBAUM, Zucker Hillside Hospital, Psychiatry Research, USA
GREGORY D. ABOWD, Northeastern University, USA and Georgia Institute of Technology, USA
MUNMUN DE CHOUDHURY, Georgia Institute of Technology, USA

In the mental health domain, patient engagement is key to designing human-centered technologies. CSCW and HCI researchers have delved into various facets of collaboration in AI research; however, previous research neglects the individuals who both produce the data and will be most impacted by the resulting technologies, such as patients. This study examines how interdisciplinary researchers and mental health patients who donate their data for AI research collaborate and how we can improve human-centeredness in mental health AI research. We interviewed patient participants, AI researchers, and clinical researchers in a federally funded mental health AI research project. We used the concept of boundary objects to understand stakeholder collaboration. Our findings reveal that the social media data provided by patient participants functioned as boundary objects that facilitated stakeholder collaboration. Although the collaboration appeared to be successful, we argue that building consensus, or understanding each other's perspectives, can improve the human-centeredness of mental health AI research. Based on the findings, we provide suggestions for human-centered mental health AI research, working with data donors as domain experts, making invisible work visible, and privacy implications.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**; • **Empirical studies in HCI**;

Additional Key Words and Phrases: human-AI interaction, AI research, mental health, boundary objects, collaboration, patient-generated data, social media

ACM Reference Format:
Dong Whi Yoo, Hayoung Woo, Sachin R. Pendse, Nathaniel Young Lu, Michael L. Birnbaum, Gregory D. Abowd, and Munmun De Choudhury. 2024. Missed Opportunities for Human-Centered AI Research: Understanding Stakeholder Collaboration in Mental Health AI Research. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 95 (April 2024), 24 pages. https://doi.org/10.1145/3637372

Authors' addresses: Dong Whi Yoo, dyoo@kent.edu, Kent State University, Kent, Ohio, USA, 44242; Hayoung Woo, Georgia Institute of Technology, Atlanta, USA, hwoo46@gatech.edu; Sachin R. Pendse, Georgia Institute of Technology, Atlanta, USA; Nathaniel Young Lu, Zucker Hillside Hospital, Psychiatry Research, Glen Oaks, New York, USA; Michael L. Birnbaum, Zucker Hillside Hospital, Psychiatry Research, Glen Oaks, New York, USA; Gregory D. Abowd, Northeastern University, Boston, USA and Georgia Institute of Technology, Atlanta, USA; Munmun De Choudhury, Georgia Institute of Technology, Atlanta, USA.

ACM 2573-0142/2024/4-ART95
https://doi.org/10.1145/3637372

1 INTRODUCTION

Artificial intelligence (AI) technologies, which utilize unprecedented amounts of data and computational power, are expected to address clinical challenges that have been unsolvable with traditional technologies and human efforts. However, the structure of data and the logic of computational approaches in AI technologies are entangled with existing societal issues, such as bias and marginalization. Scheuerman et al. [74] rightly said that the data used to build computer vision algorithms often have implicit politics embedded in them, whether it is in their collection, curation, annotation, or packaging. Many researchers have also suggested that AI technologies may even accelerate or perpetuate such problems [26, 64]. One way to understand and address the ethical issues associated with AI technologies is to examine how people who develop AI technologies interact with data and collaborate with other stakeholders. As Muller et al. pointed out, AI research is an amalgamation of human labor [55], and the resulting technologies will reflect the values and biases of the people behind the development.

Muller and Strohmayer [56] recently noted that there is emerging research in computer-supported cooperative work (CSCW) and human–computer interaction (HCI) that aims "to de-center data as the primary object of interest and to re-center data work on people and their relationships to one another through data." Such a re-centering on people, in many cases, means delving into stakeholder collaboration in AI research as the design and use of AI technologies touch upon the work, practices, and lives of many people [55]. Previous studies have explored how AI researchers work [55] and collaborate with domain experts who are not well versed in computer science and AI [48, 91]. Scholars have advocated for stakeholder collaboration in domains as diverse as online sexual harassment detection [43, 65] and public services [41, 71]. While these studies have deepened our understanding of collaboration in AI research, they either limit themselves to non-human data (e.g., in natural science research [58]) or sometimes neglect the real people behind the data.

In the domain of mental health, Chancellor et al. [16] found that data contributors are often represented as mere numbers to be entered into AI model training, while Ernala et al. [29] uncovered threats to validity that may result from disregarding data creators in AI model development. This dehumanization may not only jeopardize the reproducibility of research results [16] but also remove context from data, a key attribute of technology design [27]. Most importantly, in addition to the negative consequences of scientific and technological design efforts, dehumanization may harm the individuals whom we seek to help by increasing the stigma they face as people suffering from mental health disorders [16]. Therefore, we see an opportunity to explore collaborations between data donors and interdisciplinary researchers in AI research with a focus on human-centeredness. Human-centeredness involves treating the human subject as a person, not a data point [16], and gathering the voices and concerns of stakeholders during the collaboration [46].

The concept of boundary objects [81], which has been widely used to explore collaboration in technology design in CSCW and HCI [1, 14, 23], can be useful for understanding human-centeredness in AI research. Boundary objects originated when Star sought to explain the successful collaboration without consensus among stakeholders in a natural science project [81]. We posit that consensus is not necessary for collaboration, but it can enhance human-centeredness in terms of the mutual understanding of each other's perspectives. We use the concept of boundary objects to analyze how stakeholders in a mental health AI project collaborate without consensus and how we can increase human-centeredness by establishing consensus. Data donors in AI research projects are important not only because they are human subjects who contribute their own perspectives to the research process but also because they are the people who will be most affected by the resulting technologies [12].

In this paper, we examined the THRIVE project, a federally funded mental health AI project. The goal of the project was to explore the possibility of developing predictive algorithms to predict mental health conditions (schizophrenia spectrum disorders, bipolar disorder, and other psychiatric disorders) from digital traces of individuals diagnosed with these conditions. The participating medical institution involved in the project had recruited more than 300 participants who voluntarily donated their digital traces, including Google search history and data from social media platforms. The institution's clinical researchers, in collaboration with AI researchers from other industry and academic research institutions, analyzed the contributed data and presented promising predictive models for schizophrenia relapse [8]. This project exemplifies ongoing efforts within the fields of CSCW and HCI to gather patient-generated private social media data with explicit consent from participants [18]. The inclusion of human subjects in mental health AI research is imperative because any resulting technologies from mental health AI research will have a profound impact on vulnerable and historically marginalized people [15, 84]. Despite the high risks, AI technologies have the potential to address some of the challenges in current mental health practice, such as the availability of and access to adequate resources, lack of objective measures, and difficulties in timely intervention [15, 28]. To mitigate potential ethical issues and maximize the benefits of AI technologies, it is important to ensure that research projects developing mental health AI projects include and respect the voices of stakeholders [77], especially those who will be impacted by the resulting technologies [12].

From this, we seek answers to these research questions:

- **RQ1.** How do patients, AI researchers, and clinical researchers collaborate in a multidisciplinary, multi-institutional mental health AI research project?
- **RQ2.** How can we improve human-centeredness in collaboration in future mental health AI research?

To answer the above research questions, we conducted semi-structured interviews with eight patient participants, four AI researchers, and four clinical researchers from the federally funded mental health AI research project. Using boundary objects [81] as an analytical lens, we delve into our findings regarding the stakeholder collaboration, missed opportunities for human-centered AI research, and the presence of invisible work. We discuss suggestions for future human-centered AI research practices in mental health; we argue that patients, who will be impacted by the resulting technologies, should be included throughout the research process as domain experts, not simply as data contributors. Finally, future research should make invisible work visible to facilitate recognition of stakeholder values and peripheral stakeholders throughout the research process and in the resulting technologies.

Content Warning: This study presents accounts from individuals with mental health conditions, which may be triggering for some readers. Specifically, Section 5.3 includes discussion of suicidal ideation. We recommend skipping this section if a reader is uncomfortable with any form of discussion regarding suicidal ideation.

2 BACKGROUND

2.1 Stakeholder Collaboration and Human-Centered AI Research

Human–computer interaction (HCI) and computer-supported cooperative work (CSCW) researchers have argued that artificial intelligence (AI) research and other related data-driven endeavors, such as data science and machine learning, are human activities that are situated, contextualized, and value-driven [55]. Therefore, many researchers have investigated human labor in the AI research process. Some of the studies have focused on the roles, tasks, and tools of AI researchers [42, 55, 67, 85]. Building on early HCI and CSCW, these studies focus on how computer scientists and software engineers perform their roles and tasks and provide further theoretical and empirical understanding of the unique nature of AI researchers' work. Because AI research is often multidisciplinary, and many domains have begun to adopt AI-enabled approaches, recent studies have focused more on the collaboration between AI researchers and domain experts who are not AI experts [14, 48, 58, 60, 62, 91]. However, in most studies, the voices of the people who produce the data and who will be affected by the outcomes of the research are missing. We aimed to shed light on this gap of excluding people who will be impacted by the resulting technologies in AI research, and to explore potential suggestions for human-centered mental health AI research.

According to Muller et al.'s analysis, the people who created the data were disconnected from the AI researchers, although the authors claimed that "if we talk about data science (AI research) as a human activity, then, of course, we must talk about the people who do the work [55]." Related to this point, Chancellor et al. investigated how people who create data are dehumanized in the resulting papers on mental health AI and machine learning (ML) [16]. The risk of dehumanization and depersonalization in AI research and data practices can have implications not only for scientific integrity but also for individuals and societies, especially those who are historically marginalized. We build on the timely and important suggestions of Chancellor et al. not to ignore the people behind the data. Our study includes patient participants in a mental health AI research project who donated their social media data for research purposes and who will ultimately be impacted by the AI technologies resulting from the research. For example, one of the resulting technologies, a predictor of schizophrenia relapse, may impact patients (e.g., in terms of hospitalization recommendations) if the model is used in treatment settings.

We borrow Riedl's definition of human-centered AI, which is "a perspective on AI and ML that intelligent systems must be designed with an awareness that they are part of a larger system consisting of human stakeholders, such as users, operators, clients, and other people in close proximity [66]." Given this, we define human-centeredness in AI research as respecting the voices of stakeholders in the design of intelligent systems. Related to this point, CSCW and HCI researchers have recently conducted participatory design [57] studies in AI research involving vulnerable and marginalized groups of people such as youth [3], women, and LGBTQIA+ people [13, 33], nonprofit organizations [46], and the child welfare system [71]. These emerging lines of effort to engage diverse stakeholders in AI research shed light on the importance of human-centeredness in AI research and resulting AI technologies. This paper extends this conversation by demonstrating that patient participants did not have many opportunities to voice their perspectives during a mental health AI research project. This paper also suggests sustained engagement of patient participants as they will be impacted by the resulting technologies.

2.2 Boundary Objects

Ever since Star and her colleagues defined boundary objects to explain how different groups work together without consensus [9, 80, 81], CSCW and HCI researchers have utilized this concept to explore how people collaborate, especially utilizing technologies [1, 14, 23, 47, 61]. In this paper, we use the concept of boundary objects to analyze the collaboration between patient participants, AI researchers, and clinical researchers. Similar to Mol, who revealed that atherosclerosis can be seen and 'done' differently by patients, clinicians, and pathologists [51], we aimed to collect voices of the different stakeholders in a federally-funded mental health AI research project to understand what happened and what could have been improved. The concept of boundary objects helped us understand how the boundaries between the stakeholders were established and became blurred during the collaboration [51].

Among the many aspects of boundary objects, this paper focuses on these two tenets: interpretive flexibility and the transformation process. Interpretive flexibility means different groups interpret the same boundary objects in different ways. For example, in her fieldwork, Star found that hunters' camping spots were biologists' sources of data about speciation [81]. The transformation process was related to the changes in the format of boundary objects during collaboration. Boundary objects reside between different groups in a vague form. The groups will localize the objects to suit their own needs, which is not interdisciplinary. For example, professional biologists will take preserved samples from trappers and treat them as scientific samples, which may not serve other groups' needs. Groups navigate back and forth between a vague form and a localized form of the projects [80]. These two tenets informed our data analysis and resulted in one section of Findings (Section 5.1). We establish that patient social media functioned as boundary objects, and we argue that the lack of consensus resulted in missed opportunities to improve human-centeredness in mental health AI research.

2.3 Mental Health AI Research in CSCW and HCI

Utilizing an enormous amount of data and unprecedented computing power, AI technologies are expected to advance mental health practices and research in diagnosis, prognosis and treatment recommendation [28]. Previous mental health AI studies have utilized medical data [2, 31, 83], passive sensing data [25, 35, 53, 63, 69, 86, 90], and social media data [21, 22, 59, 68]. Social media data, which is also the data collected and analyzed in the research setting we focus on in this paper, has been explored at a population level (e.g., assessing population happiness [44]). Since the early 2010s, researchers have also started predicting the mental health status of individual users or individual posts for diverse types of mental health conditions, such as depression [21, 22, 54], anxiety [76], suicidal thoughts [24], eating disorders [19, 20], and psychotic disorders [8, 49].

Highly relevant to this paper, Thieme et al. conducted a systematic review of AI/ML studies from the computing and CSCW and HCI communities [82]. They found that some mental health AI research utilizes public social media data or pre-existing data sets, which may contribute to the lack of end-user voices in research practices [82]. The lack of end user voices is more exacerbated when the prospective users are patients rather than mental health providers; a good example is Hirsch et al.'s work on participatory design with therapists [40]. Centering on the perspectives of patient users, one of the rare examples is the work of Zakaria et al. that conducted interviews and surveys with students to understand how their prospective users manage their stress [90]. They used the data collected via interviews and surveys as ground truth when they evaluated their models; however, the specific findings were not reported.

Because this paper examined an AI project focusing on patients' social media data, a critical review of 75 papers regarding developing AI models that predict mental health status using social media data guided our work [18]. Specifically, the authors highlighted that many such existing research utilized proxies of mental health status, such as self-disclosures or memberships in mental health-related communities, when they identified and collected data. This can cause a lack of construct validity which can diminish the practical value of those AI models in real-world clinical settings. To alleviate this problem, the authors suggested pursuing close collaboration between AI researchers, clinical researchers, and patients who can donate their data with clinically validated mental health status information. Furthermore, in a different work, Chancellor et al. provided a taxonomy of ethical tensions in mental health AI research, namely a lack of ethics committee oversight in social media research, questions regarding the validity of such, and implications for stakeholders [17]. They concluded their analysis of existing ethics literature by calling to action interdisciplinary collaboration for mental health AI research. Specifically, they suggested a participatory algorithm design, which includes patient participants' voices through interviews, focus groups, and design workshops. Our paper is an answer to this call; we provide empirical evidence of how stakeholders collaborate in a case study and add our own suggestions for future

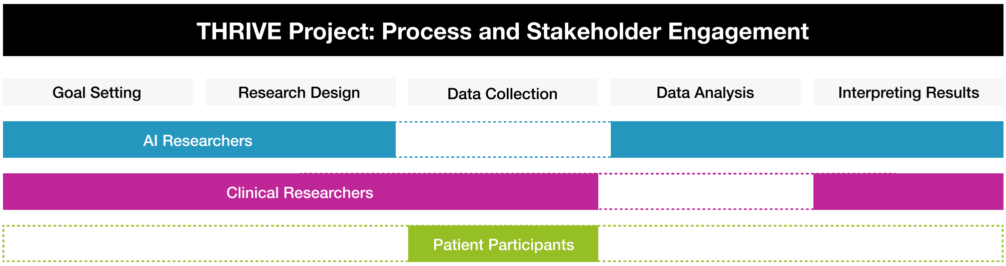

Fig. 1. The process of the THRIVE project and the extent of stakeholder involvement. The empty areas with dotted lines denote periods without stakeholder involvement. This paper involves interviews with a subset of these AI researchers, clinical researchers, and patient participants and aims to explore the dynamics of stakeholder collaboration throughout the project.

researchers who will conduct mental health AI research. In our suggestions, we expand Chancellor et al.'s participatory algorithm design by suggesting working with patients as domain experts (not as participants) and patient advisory panel approaches.

3 STUDY SETTING: THE THRIVE PROJECT

The THRIVE (Technology and Health Related Information to improVe wEllness) project is a federally funded, multi-institutional, multi-year mental health AI research project that has been collecting online traces of mental health patients, such as Google search history and social media activity, since 2018. The broader goal of the project is to better understand mental illnesses, including mood and psychotic disorders, and to find biomarkers that can identify when people need help through computational analysis of the collected data. Working closely with the relevant Institutional Review Boards (IRBs), a large psychiatric teaching hospital that is part of the research team has recruited voluntary patient participants with rigorous safeguards, such as ongoing reviews of patient well-being and consent. Inclusion criteria for the THRIVE project were individuals between the ages of 15 and 35 with a primary psychotic disorder treated at the hospital.

Clinical researchers contacted potential participants, who were patients at the psychiatric hospital, by email or in person. If the patients were interested in the process, the clinical researchers conducted intake sessions in which they verbally assessed the participants' well-being and conducted introductory onboarding sessions. The patients provided informed consent during the onboarding sessions; for patient participants under the age of 18, parental consent was obtained following participant assent. The participants then accessed their social media platforms and obtained their data archives in the form of zipped folders using the platforms' "download my archive" features. Thereafter, the participants uploaded their data archive(s) via a dedicated project website, and the uploaded data was stored on HIPAA (Health Insurance Portability and Accountability Act)-compliant secure, firewalled servers physically located and protected at the psychiatric hospital. The participants were compensated for their time and effort based on the number of social media platforms in their uploads. The participants were able to upload their data every few months if they wished to continue their participation. After data collection, the clinical researchers collaborated with AI researchers at two universities and an industry research institute to analyze the data. At the time of this writing, they were working on topics related to mental health self-disclosure [30], predicting psychotic relapse [8], understanding the emergence and characteristics of suicidal thoughts and behaviors [52], and identifying markers of schizophrenia onset to reduce

Table 1. Demographic information of patient participants.

ID	Diagnosis	Age	Sex	Race	Occupation
PP1	Schizophrenia	25-45	Male	White	Employed for wages
PP2	Bipolar disorder	18-24	Female	Asian	Student
PP3	Schizophrenia	18-24	Female	White	Student
PP4	Schizophrenia	18-24	Female	White	Student
PP5	Major depressive disorder	25-34	Female	White	Student
PP6	Schizophrenia	18-24	Female	Asian	Employed for wages
PP7	Generalized anxiety disorder	25-34	Female	American Indian or Alaska Native	Unable to work
PP8	Bipolar	24-34	Female	Black	Employed for wages

the duration of untreated psychosis [8, 38], all utilizing internet search history and social media activity voluntarily shared by consenting patient participants.

In line with the goals of this paper, to illustrate the research process and stakeholder collaboration, we have charted the overall research process as follows: *Goal Setting*, *Research Design*, *Data Collection*, *Data Analysis*, and *Interpreting Results* (Figure 1). AI researchers and clinical researchers conducted most of the process collaboratively, but *Data Collection* was led and largely conducted by the clinical researchers, while *Data Analysis* was spearheaded by the AI researchers.

4 METHODS

4.1 Participants and Interview Procedures

For this paper, the patient participants were the participants in the THRIVE project who provided their data for the project. We emailed recruitment materials to the patient participants in the project and received responses from eight participants (Table 1). We conducted a one-hour semi-structured remote interview session with these eight participants using BlueJeans and Zoom teleconferencing software. Interview questions included their motivations and expectations prior to participation, the detailed process of their participation, other people they interacted with for the project, barriers and challenges, and their opinions on future AI technologies in mental health. Patient participants received a $25 gift card for their time and effort.

Next, we recruited AI researcher participants who performed computational analyses on the online patient data in the THRIVE project (Table 2). We emailed potential participants and received responses from four AI researchers. They included one postdoctoral researcher, one Ph.D. student, and two master's students at the time of their participation in the THRIVE project. We conducted one-hour semi-structured remote interview sessions using BlueJeans and Zoom. The sessions focused on their workflow, collaboration with other stakeholders, perceived benefits and harms of the project, and their opinions on AI technologies in mental health. The AI researcher participants received no compensation.

The clinical researcher participants were affiliated with the psychiatric hospital (Table 2). There were two types of clinical researchers. The first type were the research coordinators, who were responsible for recruitment, participant assessments, onboarding sessions, and IRB protocols. At any given time, there was only one active research coordinator for the project; however, due to the multi-year nature of the project, there were a total of three different research coordinators. We were able to recruit two of them, including the first research coordinator and the current one. Second, there were two psychiatrists who were actively treating patients at the psychiatric hospital

Table 2. Demographic information of AI researchers (AR1-4) and clinical researchers (CR1-4). The occupation indicates their occupations at the time of their participation.

ID	Occupation	Age	Sex	Race
AR1	Master's student in Computer Science	25-34	Female	White
AR2	Post doctoral researcher	25-34	Male	White
AR3	Master's student in Computer Science	18-24	Male	Asian
AR4	PhD student in Computer Science	25-34	Female	Asian
CR1	Clinical research coordinator	25-34	Male	Asian
CR2	Clinical research coordinator	25-34	Female	Asian
CR3	Psychiatrist / Psychiatric researcher	35-44	Male	White
CR4	Psychiatrist / Psychiatric researcher	65+	Male	White

and who led the research efforts as domain experts. They worked on grant proposals, study design, collaboration with AI researchers and other stakeholders, and writing the resulting scientific articles. We conducted one-hour remote interviews with these participants. Semi-structured interview questions included their roles and responsibilities in the project, their expectations and workflows, perceived benefits and harms, and their opinions about future AI technologies in mental health. No compensation was given to the clinical researcher participants.

We also interviewed an IRB staff member, an IRB board member, and an information technology (IT) expert at the psychiatric hospital where patients were recruited for the THRIVE project and where data collection took place. We learned about research projects from the interviews, but because they do not directly handle patient data and AI models, we did not include them in the data analysis. However, the information we learned from these peripheral stakeholders helped us analyze our data and discuss the implications of our findings.

We recruited our participants until we found data saturation for each subgroup, meaning that new data no longer provided additional insights into the research questions [70]. All interviews were audio-recorded and transcribed with the permission of the participants. This study was approved by the lead institution's IRB.

4.2 Data Analysis

We used Braun and Clarke's method of thematic analysis [10, 11] to analyze the interview data. Certain qualitative approaches to analysis, such as grounded theory, impose a strict distinction between inductive and deductive analysis (for an excellent summary, see Muller et al. [55]). We wanted to pursue a more organic fusion of inductive and deductive analysis, so we adopted Braun and Clarke's stance on qualitative analysis [11].

The first and second authors separately coded the interview transcripts of nine participants (two patient participants, three clinical researchers, and four AI researchers). They then met and reconciled discrepancies in their codes through discussion. The first author coded the remaining seven participant interview transcripts (six patient participants and one clinical researcher). The two researchers then met and reviewed the codes to reach agreement. They generated the initial themes: tensions in motivation, frictions in the translation process, missed opportunities, and opinions about potential interventions. These initial codes were discussed among the research team members during regular research meetings.

During our iterative analysis, we noticed similarities between the patterns in our data and the concept of boundary objects [81]. We then revisited our codes and refined the themes using boundary objects as an analytical lens. In particular, a subset of the key tenets of boundary objects, such as interpretive flexibility, dynamic process, and invisible work (detailed in Section 2.2), guided the refinement of our themes. Specifically, we revised our initial themes, tensions in motivation and frictions in the translation process, into interpretive flexibility and data transformation (Section 5.1) and invisible work (Section 5.3). We also merged two initial themes, missed opportunities and opinions on potential interventions (now Section 5.4). During iteration, we decided to add the theme of building a shared vocabulary (Section 5.2) to represent the close relationship between clinical researchers and AI researchers, which ultimately underscores the missed opportunities of working with patient participants, the main argument of our data analysis.

4.3 Positionality

In this research, we follow constructionism, which means that we acknowledge our own biases and subjectivity as real people observing and interpreting the objects we study (as opposed to neutral "scientists" discovering existing objective knowledge) [4]. This is a pertinent point because our team includes HCI researchers, mental health clinicians, and medical researchers. Our team also has a diverse demographic and cultural background, including people of color, LGBTQ+ people, and immigrants. Notably, one of the authors has lived with a mood disorder for more than 15 years, which served as part of the motivation to improve the human-centeredness of mental health AI. Our personal motivations also dovetailed with our professional experiences in mental health research and practice. Collectively, the diversity of the team members' professional and personal experiences allowed us to examine stakeholder collaboration in the THRIVE project from multiple perspectives.

5 FINDINGS

5.1 Interpretive Flexibility and Transformation of Data

During the collaboration, stakeholders were able to collaborate by sending and receiving different forms of patient social media data. In this section, we follow the trajectory of the transformation of the patient social media data from the archive downloaded from the social media platforms to the result of the predictive models.

The journey of the social media data began when the patient participants requested their archives from their social media platforms. The patient participants mentioned that they were aware of the existence of their online behavior ("[Social media companies] already have our data (PP2)"), but they had not acquired or explored their own data archives prior to the study. For the patient participants, the social media data was a neglected personal history. They were also well aware of data ownership and the changes in ownership that occurred when they donated their data for the research. It seemed that they were able to claim ownership of these traces of their online behavior by donating the data for research because this donation was the very first decision they had explicitly made about the data:

> *"So I choose now with this project that I have the option to choose who also can see my information and how I use my social media." – Patient Participant (PP) 7*

For the AI researchers, the patients' social media data was a key component in improving model performance. The first part of their job was to clean this data to ensure a consistent data structure throughout the project. This task was complicated by the changing structure of the archive and the lack of direct involvement from the social media platform developers. After ensuring compatibility, the data was incorporated into the data pool and analyzed using a variety of methods. The results (usually in the form of performance metrics) were then shared with the clinical researchers. In other words, the AI researchers transformed the patients' social media data into quantifiable metrics that captured the results and performance of the AI models. In particular, they often needed more data points to improve the performance of their models on sensitive tasks, such as predicting psychotic relapse.

> *"My problem here is just trying to get more data. [...] in order to do that you just need more data right? just more data."* –AI Researcher (AR) 3

As Star explained [80], the AI researchers navigated back and forth between the vague form of a boundary object (the patients' social media archives) and the localized form of a boundary object (the results of the data analysis). It was an iterative process because they had to try different features to improve the performance of their models. The inconsistency of the data structure, as mentioned earlier, was another reason why the AI researchers had to go back and forth between the archive and their results.

In contrast, the clinical researchers did not directly interact with the patients' social media data. Although the patients' social media data was physically stored on the hospital's HIPAA-compliant servers, and the clinical researchers were primarily responsible for storing the data, they did not directly work with the data. For the clinical researchers, social media data is potential health information, and they attempted to link the AI researchers' findings to specific health information. The difference between the AI researchers and the clinical researchers in interpreting the results was that the AI researchers valued the overall performance of the aggregated data (total participants), while the clinical researchers valued the potential health information specific to individual patients. For example, while the AI researchers were discussing different analytical techniques and the corresponding performance of the entire data set, the clinical researchers were envisioning how predictive modeling could be used to treat a specific patient. However, this transformation from statistical results to an actionable treatment recommendation was not complete; the clinical researchers saw this as a future direction:

> *"Often we're looking at group data and we see you know, that in a large group of patients, we can identify differences or changes, but I think the question always, what is our ability for a given individual, you know, to identify a threshold at which we need to intervene? That's not so easy."* –Clinical Researcher (CR) 4

Moreover, this transformation of patient social media data —from latent personal history to performance metrics to (potential) health information— implies that different stakeholder groups shared a global data format (the archive of patient social media data), but had their own localized versions and representations of the shared object (Table 3). To summarize, in this section we explored how patients' social media data functioned as boundary objects between the stakeholders. The stakeholders had their own interpretation of the boundary object (interpretive flexibility), and they changed the format of the object to accomplish their tasks (transformation of data). This also shows that patient social media data as a boundary object mediated successful collaboration [81].

5.2 Building Shared Vocabulary

We found that the AI researchers and clinical researchers collaborated from the very beginning of the THRIVE project. The goal of the project (e.g., predicting relapse in schizophrenia) was often set by the clinical researchers based on their clinical expertise, but *Research Design* (Figure 1), including setting research questions, was done collaboratively. After *Research Design*, *Data Collection* was entirely led by the clinical researchers, while *Data Analysis* was mainly done by the AI researchers. However, after these steps, they collaborated on the interpretation of the results (*Interpreting Results*). To do this, the AI researchers reported having to translate the results into an easily readable format for the clinical researchers, who were not trained in computer science or computational approaches.

> *"I do remember spending quite a bit of time just trying to get things in like a workable, human readable format. [...] You need to get yourself to an end result of something that is*

Table 3. A summary of stakeholder collaboration for the concept of boundary objects.

	Patient Participants	AI Researchers	Clinical Researchers
Common goal	• Improving mental health (same for all stakeholders)		
Boundary objects	• Patients' social media data (same for all stakeholders)		
Goals	• Helping peer patients	• Contributing to computer science	• Improving individual mental health outcome
Interpretive Flexibility	• Latent personal history	• Ingredients for AI models	• Potential health information
Transformation process	-	• Transforming archives to performance metrics	• Transforming performance metrics to medical knowledge
Invisible work	• Downloading archives from social media platforms is time-consuming	• Processing patients' social media data is emotionally burdensome	• Working with the treatment team is invisible to the patients

> *interpretable by people who don't necessarily have like a super deep technical background."*
> –AI Researcher (AR) 1

During this translation phase, the AI researchers migrated their results from tools specific to data analysis (e.g., Jupyter Notebook [37]) to tools used by clinical researchers (e.g., Microsoft Excel), which required some time to become familiar with. According to our informants, although they knew how to use spreadsheets, they felt less adept and confident using them as a tool to discuss their findings; they would have used other methods had they collaborated with other AI researchers. Related research on AI research tools and collaboration [91] did not find this friction in using non-AI research-specific tools to collaborate with non-AI researchers. We believe that this point can provide important insights for the design of AI research tools. For example, user experience (UX) design tools (e.g., figma.com) have incorporated web browser-based user interfaces to allow non-designer collaborators to easily access design files. Similarly, we can envision sharing features of those AI research tools that automatically generate results in human-readable formats that non-technical researchers can easily access.

AI Researcher (AR) 4 described this close collaboration as "building shared vocabulary." We identified that this process of building a shared vocabulary consisted of two distinct phases: learning and translating. Both the clinical researchers and the AI researchers acknowledged a lack of knowledge about the other side's domain, and they spent time learning more about the domain expertise ("learning curve (CR3)" and "a good amount of learning (came) from each other (AR4)." The other phase of vocabulary building is translating the result. This happened during *Interpreting Results* (Figure 1) when they had the results from *Data Analysis*. Translating is a collaborative process where the AI researchers first explain the results, such as the performance metrics of the results, and the clinical researchers map the results to clinically meaningful information. Specifically, the AI researchers needed to explain the performance metrics of their AI models, such as area under the curve (AUC), because these metrics are not used in the clinical setting [18]. The clinical researchers also wanted to learn the acceptable range of performance metrics in previous studies. After reaching consensus on the results of their analysis, the clinical researchers mapped the results to clinical practice. At this stage, they began to see the gaps between the computational analysis and clinical practice. These gaps arose in part from the operationalization of subjective mental health symptoms. For example, they categorized each month as either a period of relative health or a period of relative illness based on hospitalizations. However, this binary distinction may not fully reflect real-world illness trajectories, which are more likely to exist along a spectrum between illness and health. Clinical researchers raised these issues during the research meetings, and both clinical and AI researchers concluded that the operationalization of symptoms was a limitation of their current approach and mentioned it as such in their resulting papers.

While the AI researchers and the clinical researchers had close interactions to build a common vocabulary, the patient participants had limited interactions with them. The patient participants only interacted with the clinical research coordinators (CR1 and CR2 in Table 2). The clinical research coordinators perform several tasks with the patient participants: recruitment, informed consent, initial interview, data sharing (participants' social media data), and iterative interviews to check in with participants. While some of the patient participants mentioned that their interactions with the clinical research coordinators were positive ("He is a person who has done everything very professionally (PP8)"), the patient participants did not have any further interactions with the research team. Some of the patient participants mentioned that they would like to learn more about how their social media data contributed to new knowledge. We suspect that this is in part due to their philanthropic motivation. For example, PP7 mentioned that she would like to know the results of the research and that she is part of the research.

> *"I was interested in learning a little bit more about, you know, why they do these projects and how these projects are kind of done. So I opened up an Instagram I used it and then you know, I put in my two cents into the research to maybe get a little bit of information back because at the end of the day, it's just an exchange of information, right? [...] I haven't gotten information back from my participation. Like I don't get information back of, of the research itself, but I know that at some point, later on, I'm going to see the THRIVE project has came up with this and this data, and I know that I'm going to be part of that."* –Patient Participant (PP) 7

Regarding the lack of interaction between the research team and the patient participants, Patient Participant 1 suggested that the project could be improved with more conversations between the researchers and the patient participants. His comment powerfully implies the lack of interaction as well as his eagerness to become more involved in the project.

> *"[The project needs more] the connectivity between the people that are doing the research and us, the I mean, this interview right here. It's it's not even an interview, it's conversation, but it itself is like it's half an hour of your time as a researcher half an hour as the as a patient. [...] So this is the kind of activity that we do need."* –Patient Participant (PP) 1

In short, AI and clinical researchers collaborated closely, which was explained as "developing a shared vocabulary" while patient participants were involved in the project only as data contributors. In the following subsections, we discuss how patient participants did not collaborate closely with other stakeholders.

5.3 Invisible Work

From the interviews we found different layers of invisible work in collaboration. There are two types of invisible work: the first includes work that is invisible to the audience of the resulting papers (things that are not reported in academic papers), and the second includes work that is invisible to other stakeholders.

When the patient participants donated their social media data to the research team, they had to download their own archives to their local devices and upload the archive to the hospital server. Initially, the research team developed an automated version so that the patient data could be transferred directly from the social media platforms to the hospital server whenever the participants logged in with their IDs and gave the research team permission. However, during the course of the THRIVE project, the social media platforms changed their application programming interfaces (APIs) polices to restrict automated app-based data access and data sharing with third parties, largely due to the Cambridge Analytica scandal [36]. The manual process of having patients download their own archive, which was the only viable alternative given the API changes, was not streamlined. On some platforms, patients had to wait a few days after requesting an archive to receive a download link. They mentioned that their busy schedules and some of their symptoms (e.g., poor concentration) interfered with their intent to download and upload data.

> *"It was kind of frustrating because sometimes the links to download the data were only valid for a few days at a time. And one of the challenges of having a psychotic disorder is having cognitive symptoms, as well as you know, negative symptoms. And oftentimes, if I was downloading data I would forget to access the link after I was done downloading, and then will expire and kind of have to start all over again.* – Patient Participant (PP) 1

We note that the patient participants were willing to go through these cumbersome data sharing procedures because they valued helping peer mental health patients. However, this value was not apparent to people outside the project because these efforts themselves were invisible.

From the perspectives of AI researchers, the data cleaning part was invisible to the audience. Previous studies have already mentioned some of the invisible labors of AI researchers related to data cleaning and processing [55]. What we found in addtion was that because it was sensitive and personal data, it was *"really heavy (AR1)"* and sometimes *"felt weird (AR3)"* for AI researchers to go through some of the content.

> *"It really sucks to have that 10,000 foot view but also to know that that's a real person who was actually suffering I think it depends on your own ability to deal with that. Your own ability to take a step back to disconnect yourself to you know, dedicate time to self care, take breaks, whatever it is that personally works for you."* –AI Researcher (AR) 1

One area of invisible work for the clinical researchers involved was the year and a half it took to get IT and legal approval from the hospital and then to set up a secure server for the study, which was not reported in the resulting papers:

> *"Getting IT approval and legal approval was a hurdle. And I think it's just because [Hospital Name] is incredibly conservative and careful when it comes to PHI (Protected Health Information). And this is sort of a complicated novel data set. So it took them a very, very, very long time to get this off the ground to develop the website. To get approval for the APIs, it was just a year and a half. Just it was a, it was a really challenging process."* –Clinical Researcher (CR) 3

This invisible work of clinical researchers demonstrated their commitment to patient safety. Clinical researchers had to delay their research progress to maintain a high level of patient safety, which was not reported in their results. We also found that this value of patient safety was invisible not only to the academic audience but also to the patient participants. Because the patient participants were all patients of the same hospital, the clinical researchers were able to reach out to their clinicians if they learned anything concerning about the patients during the onboarding process or regular check-ups.

> *"This particular patient had actually seen their doctor and we were seeing them after the doctor's visit. And while we were interviewing her, she started disclosing certain things during the assessment, you know, the staff, our rater, called me and said 'I think this person I'm not comfortable sending this person home. She's disclosing a lot of suicidal things. Wanting to kill herself, like you know, after leaving here, walking into traffic or something', and I immediately, while this patient was still sitting with the rater, I ran to the clinician, the nurse practitioner, who had seen her, and sent the patient back to the nurse practitioner, and she ended up admitting her after a bit of more conversation."* –Clinical Researcher (CR) 2

Related to this point, Chancellor and De Choudhury posed a question to the HCI community about how we can respect the physician-patient relationship and a physician's duty to treat when a research team collects and analyzes social media data that can potentially reveal an individual's negative mental health status [18]. Our findings show that careful collaboration between research and treatment teams can mitigate these concerns. However, most of our patient informants indicated that they thought the treatment team was not collaborating with the research team.

> *"She (Psychiatrist in the treatment team) is working with [Hospital] just for psychiatry, so I don't think it's related. I don't think she knows."* –Patient Participant (PP) 2

To sum, we have found that the value of patient participants in helping their peers and the value of clinical researchers in patient safety are less visible than they should be. In terms of patient engagement and human-centered AI research, we believe that patient participants should be appropriately recognized for their dedicated work and that clinical researchers should openly communicate the patient safety measures.

5.4 Missed Opportunities

While the AI researchers and clinical researchers maintained their partnership throughout the research process, the patient participants had limited interaction with the other stakeholder groups. The patient participants were contacted by the clinical researchers either in person or by email at the time of recruitment. They went through a consent and onboarding process that was also led by the clinical researchers. The consent and onboarding process was designed to provide participants with enough information to make informed decisions regarding participation. However, even though the procedures were approved by the hospital's IRB and conducted under the careful oversight of the clinical researchers, our interviews revealed that the patient participants had a limited understanding of the project. Most patient participants explicitly stated that they did not have a clear understanding of the purpose of the underlying research: *"I just don't fully ever understand the purpose of these studies (PP6)."* Furthermore, and probably as a result, some participants developed their own mental models of the project, which were slightly different from what the actual research was trying to achieve:

> *"[I thought] it's about using social media data, volunteered by individuals to see how it impacts mental health."* –Patient Participant (PP) 2

> *"The whole project I would think is to unveil some type of a correlation between people's psyche and people's mannerisms and things that are coming along with all these technologies that are advancing."* –Patient Participant (PP) 7

As mentioned in Section 3, the main purpose of the THRIVE project was to analyze patients' social media data to find early signs of mental illness. However, as the quotes above suggest, many patient participants thought the project was aimed at understanding how social media affects mental health or how social media and mental health correlate. Broadly speaking, this is not entirely incorrect, as the project could use correlations between social media activity and mental health to find potential early signals. The difference is that understanding correlations can increase knowledge, while identifying early signals can help with interventions to predict relapse. We conjecture that this limited understanding and indifference stems from their abstract motivation and trust in the research and the treatment team. Most participants mentioned that, in addition to compensation, they participated in the study to help other peer patients. That is, as long as the research could help others, they did not care about the details.

> *"I didn't inquire. I didn't ask. But just that it would be used for research. So that's all that matters."* –Patient Participants (PP) 5

We believe that this limited understanding of the project is one of the missed opportunities to improve patient engagement, because understanding the purpose is the starting point for participation and engagement [15]. This limited understanding could lead to unwanted consequences; some participants might not have donated their data if they understood the actual purpose of the study; some non-participants might have participated if the research purpose had been better communicated.

The patient participants had their own perspectives and opinions regarding potential future technologies based on this research, which were not heard during the project. During the interview, we briefly introduced the potential technologies that can provide algorithmic predictions. The patient participants were ambivalent about them. One of the participants who was positive about the potential intervention, PP4, mentioned that when she had some early signs of psychotic disorder, such as severe anxiety and insomnia, her parents did not believe that she was experiencing symptoms, which delayed the start of treatment for her. So she felt that "it would have been helpful" for her because she "just didn't have the words for (her disorder)" even though she knew she was experiencing signs of mental illness (PP4).

Although some participants acknowledged that the technologies could be helpful, many patient participants emphasized discretion regarding potential interventions. Some participants mentioned that some people might be uncomfortable with algorithmic predictions regarding their psychosis: "I personally support the mission so I don't care too much. But the others may feel uncomfortable with that (PP2)." Our informants further explained the reasons for their discreet approach to these technologies: stigma and power dynamics. PP3 mentioned that "I should be notified first (about the predictions) before other people are notified" because of the sensitive nature of the information. PP1 further mentioned that predicting psychosis can be "very invasive and potentially traumatizing." They wanted to share the algorithmic predictions with their treatment team; however, they emphasized the importance of consent due to the power dynamics between their clinicians and themselves:

> *"I think with doctors and like the psychiatrist and stuff, I think there's a little bit of like, as if they have like authority over the patients and I think we deserve to have our own say despite any mental illness. I don't really I wouldn't want them to know. I wouldn't want that information being sent to my doctors without my without my consent."* –Patient Participant (PP) 5

We believe that this call for discretion reminds us of the socio-technical aspects of mental health AI technologies. While the research team and the resulting papers focused on technical accuracy and performance, there are many social challenges that mental health AI technologies must consider, such as consent beyond data collection, the therapeutic alliance, and patient participation.

We emphasize that these important and insightful patient perspectives went unheard and unseen throughout the research project. However, the AI researchers did acknowledge the people behind the data emphasizing that "the research wouldn't be possible without their contributions (AR4)." At the same time, they did not view the patient participants as their collaborators ("So they're not really working with me to, like, you know, understanding the data (AR3)"). The patient participants had a dehumanized perception of the AI researchers and assumed that the AI researchers also saw them as nothing more than another data point. Some of them thought that there was no human labor involved in the data analysis ("I was being told that it's mostly like an AI doing everything (PP4)"), or even if there was a human researcher, they speculated that the human researchers would view their data as nothing more than numbers.

> *"I'm just another name and a number to them. I don't think they care."* — Patient Participant (PP) 3

We believe that patients' limited understanding of the project and their unheard perspectives on future technologies are missed opportunities to improve patient engagement and human-centeredness. As AR4 stated, "If it's done with care and concern for participants and if it's done ethically, it is done with all stakeholders in mind and working together." Patient engagement should be considered more deeply and rigorously in future research.

6 DISCUSSION

6.1 Patient-Contributed Social Media Data as Boundary Objects

The goal of our study was to understand how different stakeholders collaborate and how their perspectives are reflected in the developed prediction algorithms. Our findings show that the patient participants, the AI researchers, and the clinical researchers had a common goal of developing new technologies for mental health, and the partnership between the clinical researchers and the AI researchers enabled them to incorporate their own perspectives into the resulting technologies. However, as our findings revealed, the patient participants did not have a good understanding of the project, they did not have the opportunity to interact directly with the AI researchers, and they did not have the opportunity to voice their opinions in the overall research process. This disconnect in a seemingly successful collaboration is why we utilized the concept of boundary objects. When Star coined the term boundary object, one of her motivations was to explain (or understand) how members of different communities of practice were able to work without consensus. The concept of consensus motivated Star to take a close look at collaboration because what she learned from her field study was that different communities of practice were able to collaborate successfully without consensus [80, 81]. From our findings, we were able to narrow down the meaning of consensus to two different facets: the first is understanding the perspectives of other stakeholders, and the second is the ability to express opinions during collaboration.

The long-term team-building and learning process of this project enabled the clinical researchers and the AI researchers to understand each other's perspectives. On the other hand, the patient participants did not have the opportunity to develop a close relationship with the AI researchers or the clinical researchers, even though some of them mentioned that they wanted to learn more about the research, especially the results of the project. Similarly, the patient participants did not have opportunities to voice their thoughts and opinions during their participation in the THRIVE project, even though they clearly articulated their concerns regarding the potential resulting technologies during this interview study. In other words, during the research process, there was almost no consensus from the patient participants' perspective despite the overall successful academic collaboration. We understood that this collaboration without consensus was possible thanks to the social media data contributed by the patients, which functioned as a boundary object.

More specifically, the procedures and protocols surrounding the social media data functioned as boundary objects.

Our findings showed that the social media data as boundary objects in our case study had interpretive flexibility, which means that different stakeholders have their own interpretations of boundary objects [80, 81]. The patient participants considered it as an (almost hidden) personal history that contains various aspects of their daily lives. For the AI researchers, this concept of personal history changed to the ingredients for developing AI models. Even though our informants mentioned that they were aware of the existence of the real people behind the data, the nature of their work involved converting the complex and messy trajectory of online behaviors into well-structured and organized formulas and numbers that could be evaluated by the performance metrics of AI models. Finally, our clinical researchers considered the patient social media data as clinical information relevant to the patient's condition. In this process, the meaning of the numbers generated by the AI models and tuned by the AI researchers was transformed again, this time into health information, which is another level of abstraction. As Chancellor et al. poignantly pointed out, because of the postpositivist approaches to AI research, it is very easy to forget the "human" in the research process [16].

In addition to bettering our understanding of collaboration and human-centeredness, thinking about data as boundary objects can deepen our understanding of data itself. Muller et al. provided insightful categories of data in AI research, focusing on how AI researchers perceive data: "data as given, as captured, as curated, as designed, and as created [55]." Our findings extend these categories of data to the particular forms of data that shape collaboration in AI research. In our case study, the data were not granted or created by researchers. The data was shared by patient participants, with an abstract belief that their donation could help others. The AI researchers were aware of the sensitive and personal nature of the data when they transformed the data into aggregated performance metrics. For clinical researchers, social media data was another window into understanding the health of individual patients.

Finally, we would like to contribute to the concept of boundary objects by positing that, even though consensus is not necessary for successful collaboration, consensus can enhance the human-centeredness of collaboration. When we consider the example of the Natural Science Museum of Star's work [81] – human-centeredness – the voices of the people involved in the research – was not particularly important because the goal of the project was to preserve the nature of the area; it did not directly affect the people in the community. However, in our case study, the development of new technologies for mental health can have multiple and significant impacts on the people in the community of interest. Lee et al. [45] spoke elaborately about the multi-faceted ways in which AI-driven and AI-mediated technologies carry the potential for both help and harm. This is particularly relevant given that the potential targets and beneficiaries of the resulting technologies are schizophrenia patients within the THRIVE project, who must take medications that can cause severe side effects or be hospitalized for intensive care. As such, the predictions and recommendations of potential AI technologies will be highly influential [88, 89]. Therefore, we argue that research collaborations to develop AI models that are intended to intervene in human behavior in high-stakes domains should consider facilitating consensus between the various groups of stakeholders. Our findings also highlight the fact that the voices of the people who will be affected by AI research may be muted during the research process. We further discuss suggestions for improving stakeholder collaboration in the next subsection.

6.2 Implications for Human-Centered AI Research

6.2.1 Working with Data Donors as Domain Experts. The fundamental concept of human-centered AI is the engagement of stakeholders during the AI development process [5, 66]. In the case of

AI research projects, this engagement may include the development of research questions and data analysis. In the THRIVE project, both the clinical researchers and the AI researchers were deeply involved in the process; however, the patient participants were not given opportunities to be more involved. One of the reasons for this is that the patient participants and the AI researchers did not have channels for collaboration. As our findings suggest, the AI researchers would have felt uncomfortable working directly with the patient participants due to a lack of training. Even the clinical researchers and the patient participants spent most of their time and effort on data collection and patient safety rather than discussing future technologies.

As discussed in Section 5.4, our patient participants have their own perspectives and valid concerns regarding future technologies that the AI researchers would like to build. In addition, some of the patient participants mentioned their interest in long-term participation, such as learning the results of the project. Therefore, in line with action research and participatory human-centered AI approaches [3, 13, 33, 39, 46, 72], we argue that AI researchers should work with patients as domain experts, similar to how AI researchers work closely with clinicians, to better understand problem spaces, design data analysis plans, and interpret results. At first glance, patients' experiences and knowledge may not be systematic or rigorous enough to inform AI research; however, there are pioneering studies working with community members ("non-experts") to improve AI models [13, 32, 46, 72]. In addition to improving AI models, patient input can guide the implementation process by providing an ethical framework from the patient's perspectives. We argue that future human-centered mental health AI research should work with patient participants as domain experts to not only improve the performance of AI models but also to better envision future technologies that can address patient concerns.

6.2.2 Making Invisible Work Visible. Another implication of our findings is to make invisible work visible. In Findings, we reported the invisible work of stakeholders: the data sharing process of patients, the emotional burden of AI researchers, and the patient safety measures of clinical researchers. By increasing the visibility of this backstage work, we can improve human-centeredness in complex sociotechnical systems [73]. Specifically, we envision that through this we will achieve two benefits: recognizing peripheral stakeholders and soliciting stakeholder values.

First, we can make the process of sharing patient data more explicit in the resulting academic papers. As we have seen in the THRIVE project, in the current resulting academic papers, the commitment to data sharing is simply mentioned as patients uploading their archives to the research website. This statement not only diminishes the time and effort of patients but also excludes other peripheral stakeholders in this collaboration: the social media platforms (or the manufacturers of the devices used to collect the patients' data). The decisions of these industry stakeholders affect AI research efforts; research teams change their data collection methods when companies change API policies or the data structures of user archives. It may not simply be additional work that research teams have to do; the scope of projects and collaboration is limited by the types of data these stakeholders allow their users to download and donate to research efforts. We envision that elaborating on the process of participant data sharing will clarify the values of patient participants in supporting peer patients, as well as how these data infrastructures affect scientific efforts.

Another type of invisible work we identified, the emotional burden of AI researchers, may be an opportunity to elicit stakeholder values for improving collaboration and making the resulting technologies ethical. The emotional burden we identified among our AI researcher participants was closely related to the ethical responsibilities of AI researchers, as AR1 mentioned that she thought empathy for patients was important in this high-stakes AI research. As researchers explore high-stakes application areas, we may need to document our perspectives and responsibilities as we interact with people's data and potential AI models that may impact those people. Recent efforts to document the ethical aspects of AI models (e.g., consumer labels [75] and model cards [50]) may consider these "ethical statements from the research team" to be included in this documentation.

6.2.3 Implications for Privacy. AI research that utilizes individuals' personal data has potential privacy issues because AI models can use massive amounts of data in invisible ways [79]. HCI researchers have pointed out privacy concerns and potential solutions for the use of personal data [23, 34, 78, 87]. In our study, the patient participants did not express specific privacy concerns regarding donating their social media archives. Some of them mentioned that they were concerned before the onboarding sessions, but after learning about the THRIVE project's safeguards, they felt comfortable sharing their data. However, this does not mean that privacy issues are not important in human-centered mental health AI research. We would like to highlight the importance of considering non-participant perspectives in human-centered AI research, particularly in relation to privacy concerns.

We conjectured that our patient informants were not very concerned about privacy because they had positive impressions regarding their treatment and the hospital. They also explicitly mentioned that they trusted the research team regarding the use of the social media data provided by the patients. However, we must keep in mind that these participants were recruited through the hospital, so it may be that individuals who already trust the hospital decided to participate in the study. In other words, individuals who do not trust the hospital or who are marginalized by the healthcare system may decide not to participate in the study. HCI and science, technology, and society (STS) researchers have argued that the perspectives of non-users, people who choose not to use a technology or who are marginalized by the technology, have provided insights and design implications for improving the human-centeredness of technologies [6, 7]. Similarly, we argue that we need to solicit the voices of individuals who did not choose to participate in the AI study due to trust and/or privacy issues.

6.3 Limitations

This paper is based on the qualitative investigation of a mental health AI research project conducted by research institutes in the eastern part of the United States. We note that the majority of patient participants were female. Our findings may not be generalizable to other AI research projects, such as those in non-US contexts or with different groups of participants. In this paper, we pursued transferability rather than generalizability [39] by providing a thick description of what we heard and learned, while acknowledging that we as researchers have our own biased and subjective perspectives during our research process. We also acknowledge that there are other stakeholders in AI research, such as funding agencies, institutional review boards, and institute administrators. Because we focused on collaboration mediated by patient social media data, we investigated stakeholders who work directly with the data. Future research on these other stakeholders will provide valuable insights. We also note that the potential resulting technologies that the THRIVE project is attempting to build may have a negative impact on patients. The resulting technologies may suffer from biases, such as detecting more positives in marginalized groups, or may limit patient autonomy. Our intention is to identify these risks and propose patient engagement to address them. We hope that our findings and suggestions can inspire other researchers to creatively design their own human-centered AI research practices.

7 CONCLUSION

We examined stakeholder collaboration in a federally funded mental health AI research project. In this project, patient participants donated their own social media archives for research purposes. AI researchers and clinical researchers collaborated to transform social media data into AI models and potential clinical information. Using the boundary object concept, we found that these stakeholders shared a common goal of improving mental health while also maintaining individually specific goals, such as helping peer patients for the patient participants, contributing to computer science research for the AI researchers, and improving individual mental health outcomes for the clinical researchers. These different perspectives resulted in how they viewed the patients' social media (interpretive flexibility of boundary objects); it was a donation for the patient participants, it was components for improving model performance for the AI researchers, and it was potential health information for the clinical researchers. The social media data mediated collaboration among the stakeholders, although they did not always have consensus — i.e., a deep understanding of each other. This was particularly true for the patient participants. Even though Star suggested in the concept of boundary objects that consensus is not necessary for successful collaboration, we argue that the research team would have been able to improve human-centeredness if they had facilitated consensus. One example of consensus was the close partnership between the AI researchers and the clinical researchers; they described their collaboration as a learning process and building a common vocabulary. Similarly, we envision patient participants contributing to AI research projects as domain experts at various stages of the research, such as voicing their concerns, contributing to data analysis, and envisioning future technologies. We urge future AI researchers to actively engage their data contributors, such as patient participants, in their research projects to improve the human-centeredness of AI research and resulting technologies.

ACKNOWLEDGMENTS

Yoo, Pendse, Birnbaum, and De Choudhury were partly supported through NIH grant R01MH117172 and through a cooperative research agreement between Northwell Health and Georgia Tech. De Choudhury was also partly supported through a grant from the American Foundation for Suicide Prevention.

REFERENCES

[1] Rikke Aarhus and Stinne Aaløkke Ballegaard. 2010. Negotiating boundaries: Managing disease at home. *Conference on Human Factors in Computing Systems - Proceedings* 2 (2010), 1223–1232. https://doi.org/10.1145/1753326.1753509

[2] Marios Adamou, Grigoris Antoniou, Elissavet Greasidou, Vincenzo Lagani, Paulos Charonyktakis, and Ioannis Tsamardinos. 2018. Mining free-text medical notes for suicide risk assessment. In *Proceedings of the 10th hellenic conference on artificial intelligence*. 1–8.

[3] Karla Badillo-Urquiola, Diva Smriti, Brenna McNally, Evan Golub, Elizabeth Bonsignore, and Pamela J Wisniewski. 2019. Stranger danger! social media app features co-designed with children to keep them safe online. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. 394–406.

[4] Shaowen Bardzell. 2010. Feminist HCI: Taking stock and outlining an agenda for design. *Conference on Human Factors in Computing Systems - Proceedings* 2, 1301–1310. https://doi.org/10.1145/1753326.1753521

[5] Eric PS Baumer. 2017. Toward human-centered algorithm design. *Big Data & Society* 4, 2 (2017), 2053951717718854.

[6] Eric PS Baumer. 2018. Socioeconomic Inequalities in the Non use of Facebook. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–14.

[7] Wiebe E Bijker. 1997. *Of bicycles, bakelites, and bulbs: Toward a theory of sociotechnical change*. MIT press.

[8] Michael L. Birnbaum, Sindhu Kiranmai Ernala, Asra F. Rizvi, Elizabeth Arenare, Anna R. Van Meter, Munmun De Choudhury, and John M. Kane. 2019. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. *npj Schizophrenia* 5 (12 2019), 17. Issue 1. https://doi.org/10.1038/s41537-019-0085-9

[9] Geoffrey C Bowker and Susan Leigh Star. 2000. *Sorting things out: Classification and its consequences*. MIT press.

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3 (2006), 77–101. Issue 2. https://doi.org/10.1191/1478088706QP063OA

[11] Virginia Braun and Victoria Clarke. 2016. (Mis)conceptualising themes, thematic analysis, and other problems with Fugard and Potts' (2015) sample-size tool for thematic analysis. *International Journal of Social Research Methodology* 19 (11 2016), 739–743. Issue 6. https://doi.org/10.1080/13645579.2016.1195588

[12] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[13] Caroline Bull, Hanan Aljasim, and Douglas Zytko. 2021. Designing Opportunistic Social Matching Systems for Women's Safety During Face-to-Face Social Encounters. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. 23–26.

[14] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2021. Onboarding Materials as Cross-functional Boundary Objects for Developing AI Assistants. *Conference on Human Factors in Computing Systems - Proceedings* (5 2021). https://doi.org/10.1145/3411763.3443435

[15] Sarah Carr. 2020. 'AI gone mental': engagement and ethics in data-driven technology for mental health. *https://doi.org/10.1080/09638237.2020.1714011* 29 (3 2020), 125–130. Issue 2. https://doi.org/10.1080/09638237.2020.1714011

[16] Stevie Chancellor, Eric P.S. Baumer, and Munmun De Choudhury. 2019. Who is the "human" in human-centered machine learning: The case of predicting mental health from social media. *Proceedings of the ACM on Human-Computer Interaction* 3 (2019), 32. Issue CSCW. https://doi.org/10.1145/3359249

[17] Stevie Chancellor, Michael L Birnbaum, Eric D Caine, Vincent MB Silenzio, and Munmun De Choudhury. 2019. A taxonomy of ethical tensions in inferring mental health states from social media. In *Proceedings of the conference on fairness, accountability, and transparency*. 79–88.

[18] Stevie Chancellor and Munmun De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine* 3, 1 (2020), 1–11.

[19] Stevie Chancellor, Zhiyuan Lin, Erica L Goodman, Stephanie Zerwas, and Munmun De Choudhury. 2016. Quantifying and predicting mental illness severity in online pro-eating disorder communities. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 1171–1184.

[20] Stevie Chancellor, Tanushree Mitra, and Munmun De Choudhury. 2016. Recovery amid pro-anorexia: Analysis of recovery in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2111–2123.

[21] Xuetong Chen, Martin D Sykora, Thomas W Jackson, and Suzanne Elayan. 2018. What about mood swings: Identifying depression on twitter with temporal measures of emotions. In *Companion Proceedings of the The Web Conference 2018*. 1653–1660.

[22] Munmun De Choudhury and Michael Gamon. 2013. Predicting Depression via Social Media. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media* 2, 128–137.

[23] Chia-Fang Chung, Kristin Dew, Allison M Cole, Jasmine Zia, James A Fogarty, Julie A Kientz, and Sean A Munson. 2016. Boundary Negotiating Artifacts in Personal Informatics: Patient-Provider Collaboration with Patient-Generated Data. *Proceedings of CSCW* (2016), 768–784. https://doi.org/10.1145/2818048.2819926

[24] Glen Coppersmith, Ryan Leary, Eric Whyne, and Tony Wood. 2015. Quantifying suicidal ideation via language usage on social media. In *Joint statistics meetings proceedings, statistical computing section, JSM*, Vol. 110.

[25] Orianna DeMasi and Benjamin Recht. 2017. A step towards quantifying when an algorithm can and cannot predict an individual's wellbeing. 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*. 763–771.

[26] Sunipa Dev, Masoud Monajatipoor, Anaelia Ovalle, Arjun Subramonian, Jeff M Phillips, and Kai-Wei Chang. 2021. Harms of gender exclusivity and challenges in non-binary representation in language technologies. *EMNLP* (2021).

[27] Paul Dourish. 2004. What we talk about when we talk about context. *Personal and ubiquitous computing* 8, 1 (2004), 19–30.

[28] Simon D'Alfonso. 2020. AI in mental health. *Current Opinion in Psychology* 36 (2020), 112–117.

[29] Sindhu Kiranmai Ernala, Michael L Birnbaum, Kristin A Candan, Asra F Rizvi, William A Sterling, John M Kane, and Munmun De Choudhury. 2019. Methodological gaps in predicting mental health states from social media: triangulating diagnostic signals. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–16.

[30] Sindhu Kiranmai Ernala, Tristan Labetoulle, Fred Bane, Michael L Birnbaum, Asra F Rizvi, John M Kane, and Munmun De Choudhury. 2018. Characterizing Audience Engagement and Assessing Its Impact on Social Media Disclosures of Mental Illnesses. *International AAAI Conference on Web and Social Media*. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17884

[31] Chaonan Feng, Huimin Gao, Xuefeng B Ling, Jun Ji, and Yantao Ma. 2018. Shorten bipolarity checklist for the differentiation of subtypes of bipolar disorder using machine learning. In *Proceedings of the 2018 6th International Conference on Bioinformatics and Computational Biology*. 162–166.

[32] William R Frey, Desmond U Patton, Michael B Gaskell, and Kyle A McGregor. 2020. Artificial intelligence and inclusion: Formerly gang-involved youth as domain experts for analyzing unstructured twitter data. *Social Science Computer Review* 38, 1 (2020), 42–56.

[33] Nicholas Furlo, Jacob Gleason, Karen Feun, and Douglas Zytko. 2021. Rethinking Dating Apps as Sexual Consent Apps: A New Use Case for AI-Mediated Communication. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. 53–56.

[34] Sandra Gabriele and Sonia Chiasson. 2020. Understanding fitness tracker users' security and privacy knowledge, attitudes and behaviours. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[35] Martin Gjoreski, Hristijan Gjoreski, Mitja Luštrek, and Matjaž Gams. 2016. Continuous stress detection using a wrist device: in laboratory and real life. In *proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct*. 1185–1193.

[36] Felipe González, Yihan Yu, Andrea Figueroa, Claudia López, and Cecilia Aragon. 2019. Global reactions to the cambridge analytica scandal: A cross-language social media study. In *Companion Proceedings of the 2019 world wide web conference*. 799–806.

[37] Brian Granger, Chris Colbert, and Ian Rose. 2017. JupyterLab: The next generation jupyter frontend. *JupyterCon 2017* (2017).

[38] Katrin Hänsel, Inna Wanyin Lin, Michael Sobolev, Whitney Muscat, Sabrina Yum-Chan, Munmun De Choudhury, John M Kane, and Michael L Birnbaum. 2021. Utilizing instagram data to identify usage patterns associated with schizophrenia spectrum disorders. *Frontiers in Psychiatry* 12 (2021).

[39] Gillian R. Hayes. 2011. The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction* 18 (8 2011), 15:1–15:20. Issue 3. https://doi.org/10.1145/1993060.1993065

[40] Tad Hirsch, Kritzia Merced, Shrikanth Narayanan, Zac E Imel, and David C Atkins. 2017. Designing contestability: Interaction design, machine learning, and mental health. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. 95–99.

[41] Naja Holten Møller, Irina Shklovski, and Thomas T Hildebrandt. 2020. Shifting concepts of value: Designing algorithmic decision-support systems for public services. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–12.

[42] Mary Beth Kery, Marissa Radensky, Mahima Arya, Bonnie E John, and Brad A Myers. 2018. The story in the notebook: Exploratory data science using a literate programming tool. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–11.

[43] Seunghyun Kim, Afsaneh Razi, Gianluca Stringhini, Pamela J Wisniewski, and Munmun De Choudhury. 2021. A Human-Centered Systematic Literature Review of Cyberbullying Detection Algorithms. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–34.

[44] Adam DI Kramer. 2010. An unobtrusive behavioral model of" gross national happiness". In *Proceedings of the SIGCHI conference on human factors in computing systems*. 287–290.

[45] Ellen E Lee, John Torous, Munmun De Choudhury, Colin A Depp, Sarah A Graham, Ho-Cheol Kim, Martin P Paulus, John H Krystal, and Dilip V Jeste. 2021. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 6, 9 (2021), 856–864.

[46] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, et al. 2019. WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–35.

[47] Wayne G Lutters and Mark S Ackerman. 2002. Achieving Safety: A Field Study of Boundary Objects in Aircraft Technical Support. *Proceedings of the 2002 ACM conference on Computer supported cooperative work - CSCW '02* (2002). https://doi.org/10.1145/587078

[48] Yaoli Mao, Dakuo Wang, Michael Muller, Kush R. Varshney, Ioana Baldini, Casey Dugan, and Aleksandra Mojsilovic. 2019. How data scientists work together with domain experts in scientific collaborations: To find the right answer or to ask the right qestion? *Proceedings of the ACM on Human-Computer Interaction* 3 (12 2019). Issue GROUP. https://doi.org/10.1145/3361118

[49] Margaret Mitchell, Kristy Hollingshead, and Glen Coppersmith. 2015. Quantifying the language of schizophrenia in social media. In *Proceedings of the 2nd workshop on Computational linguistics and clinical psychology: From linguistic signal to clinical reality*. 11–20.

[50] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. *FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency* (1 2019), 220–229. https://doi.org/10.1145/3287560.3287596

[51] Annemarie Mol. 2002. *The body multiple: Ontology in medical practice*. Duke University Press.

[52] Khatiya C Moon, Anna R Van Meter, Michael A Kirschenbaum, Asra Ali, John M Kane, and Michael L Birnbaum. 2021. Internet search activity of young people with mood disorders who are hospitalized for suicidal thoughts and behaviors: qualitative study of Google search activity. *JMIR mental health* 8, 10 (2021), e28262.

[53] Mehrab Bin Morshed, Koustuv Saha, Richard Li, Sidney K D'Mello, Munmun De Choudhury, Gregory D Abowd, and Thomas Plötz. 2019. Prediction of mood instability with passive sensing. *Proceedings of the ACM on Interactive, Mobile,*

Wearable and Ubiquitous Technologies 3, 3 (2019), 1–21.

[54] Danielle Mowery, Hilary Smith, Tyler Cheney, Greg Stoddard, Glen Coppersmith, Craig Bryan, Mike Conway, et al. 2017. Understanding depressive symptoms and psychosocial stressors on Twitter: a corpus-based study. *Journal of medical Internet research* 19, 2 (2017), e6895.

[55] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How data science workers work with data. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300356

[56] Michael Muller and Angelika Strohmayer. 2022. Data Silences in Data Preparation: Toward a Taxonomy. In *ACM CHI Conference on Human Factors in Computing Systems*.

[57] Michael J Muller and Sarah Kuhn. 1993. Participatory design. *Commun. ACM* 36, 6 (1993), 24–28.

[58] Andrew B. Neang, Will Sutherland, Michael W. Beach, and Charlotte P. Lee. 2021. Data Integration as Coordination: The Articulation of Data Work in an Ocean Science Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 4 (1 2021), 1–25. Issue CSCW3. https://doi.org/10.1145/3432955

[59] Albert Park, Mike Conway, and Annie T Chen. 2018. Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach. *Computers in human behavior* 78 (2018), 98–112.

[60] Samir Passi and Steven J. Jackson. 2018. Trust in data science: Collaboration, translation, and accountability in corporate data science projects. *Proceedings of the ACM on Human-Computer Interaction* 2 (11 2018). Issue CSCW. https://doi.org/10.1145/3274405

[61] Andreas F. Phelps and Madhu Reddy. 2009. The influence of boundary objects on group collaboration in construction project teams. *GROUP'09 - Proceedings of the 2009 ACM SIGCHI International Conference on Supporting Group Work* (2009), 125–128. https://doi.org/10.1145/1531674.1531693

[62] David Piorkowski, Soya Park, April Yi Wang, Dakuo Wang, Michael Muller, and Felix Portnoy. 2021. How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study. *Proceedings of the ACM on Human-Computer Interaction* 5 (4 2021). Issue CSCW1. https://doi.org/10.1145/3449205

[63] Mashfiqui Rabbi, Shahid Ali, Tanzeem Choudhury, and Ethan Berke. 2011. Passive and in-situ assessment of mental and physical well-being using mobile sensors. In *Proceedings of the 13th international conference on Ubiquitous computing*. 385–394.

[64] Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. 2020. Saving face: Investigating the ethical concerns of facial recognition auditing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. 145–151.

[65] Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Gianluca Stringhini, Thamar Solorio, Munmun De Choudhury, and Pamela J Wisniewski. 2021. A Human-Centered Systematic Literature Review of the Computational Approaches for Online Sexual Risk Detection. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–38.

[66] Mark O Riedl. 2019. Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies* 1, 1 (2019), 33–36.

[67] Adam Rule, Aurélien Tabard, and James D Hollan. 2018. Exploration and explanation in computational notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[68] Koustuv Saha and Munmun De Choudhury. 2017. Modeling stress with social media around incidents of gun violence on college campuses. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–27.

[69] Asif Salekin, Jeremy W Eberle, Jeffrey J Glenn, Bethany A Teachman, and John A Stankovic. 2018. A weakly supervised learning framework for detecting social anxiety and depression. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2, 2 (2018), 1–26.

[70] Benjamin Saunders, Julius Sim, Tom Kingstone, Shula Baker, Jackie Waterfield, Bernadette Bartlam, Heather Burroughs, and Clare Jinks. 2018. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity* 52 (2018), 1893–1907.

[71] Devansh Saxena, Karla Badillo-Urquiola, Pamela J Wisniewski, and Shion Guha. 2021. A framework of high-stakes algorithmic decision-making for the public sector developed through a case study of child-welfare. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–41.

[72] Devansh Saxena and Shion Guha. 2020. Conducting participatory design to improve algorithms in public services: Lessons and challenges. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*. 383–388.

[73] Devansh Saxena, Seh Young Moon, Dahlia Shehata, and Shion Guha. 2022. Unpacking Invisible Work Practices, Constraints, and Latent Power Relationships in Child Welfare through Casenote Analysis. In *CHI Conference on Human Factors in Computing Systems*. 1–22.

[74] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do datasets have politics? Disciplinary values in computer vision dataset development. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–37.

[75] Christin Seifert, Stefanie Scherzinger, and Lena Wiese. 2019. Towards Generating Consumer Labels for Machine Learning Models. *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*, 173–179.

[76] Judy Hanwen Shen and Frank Rudzicz. 2017. Detecting anxiety through reddit. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*. 58–65.

[77] C Estelle Smith, Bowen Yu, Anjali Srivastava, Aaron Halfaker, Loren Terveen, and Haiyi Zhu. 2020. Keeping community in the loop: Understanding wikipedia stakeholder values for machine learning-based systems. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[78] Stephen Snow, Awais Hameed Khan, Stephen Viller, Ben Matthews, Scott Heiner, James Pierce, Ewa Luger, Richard Gomer, and Dorota Filipczuk. 2020. Speculative Designs for Emergent Personal Data Trails: Signs, Signals and Signifiers. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[79] Bernd Carsten Stahl and David Wright. 2018. Ethics and privacy in AI and big data: Implementing responsible research and innovation. *IEEE Security & Privacy* 16, 3 (2018), 26–33.

[80] Susan Leigh Star. 2010. This is Not a Boundary Object: Reflections on the Origin of a Concept:. *http://dx.doi.org/10.1177/0162243910377624* 35 (8 2010), 601–617. Issue 5. https://doi.org/10.1177/0162243910377624

[81] Susan Leigh Star and James R. Griesemer. 1989. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science* 19 (1989), 387–420. Issue 3. https://doi.org/10.1177/030631289019003001

[82] Anja Thieme, Danielle Belgrave, and Gavin Doherty. 2020. Machine Learning in Mental Health. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27 (8 2020), 34. Issue 5. https://doi.org/10.1145/3398069

[83] Truyen Tran, Dinh Phung, Wei Luo, Richard Harvey, Michael Berk, and Svetha Venkatesh. 2013. An integrated framework for suicide risk prediction. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1410–1418.

[84] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–14.

[85] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing theory-driven user-centric explainable AI. *Conference on Human Factors in Computing Systems - Proceedings* (5 2019). https://doi.org/10.1145/3290605.3300831

[86] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. 3–14.

[87] Dong Whi Yoo, Aditi Bhatnagar, Sindhu Kiranmai Ernala, Asra Ali, Michael L Birnbaum, Gregory D Abowd, and Munmun De Choudhury. 2023. Discussing Social Media During Psychotherapy Consultations: Patient Narratives and Privacy Implications. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–24.

[88] Dong Whi Yoo, Michael L Birnbaum, Anna R Van Meter, Asra F Ali, Elizabeth Arenare, Gregory D Abowd, and Munmun De Choudhury. 2020. Designing a Clinician-Facing Tool for Using Insights From Patients' Social Media Activity: Iterative Co-Design Approach. *JMIR Mental Health* 7 (2020), e16969. Issue 8. https://doi.org/10.2196/16969

[89] Dong Whi Yoo, Sindhu Kiranmai Ernala, Bahador Saket, Domino Weir, Elizabeth Arenare, Asra F Ali, Anna R Van Meter, Michael L Birnbaum, Gregory D Abowd, and Munmun De Choudhury. 2021. Clinician perspectives on using computational mental health insights from patients' social media activities: design and qualitative evaluation of a prototype. *JMIR mental health* 8, 11 (2021), e25455.

[90] Camellia Zakaria, Rajesh Balan, and Youngki Lee. 2019. StressMon: scalable detection of perceived stress and depression using passive sensing of changes in work Routines and group interactions. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–29.

[91] Amy X Zhang, Michael Muller, and Dakuo Wang. 2020. How do data science workers collaborate? roles, workflows, and tools. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–23.

Received January 2023; revised July 2023; accepted November 2023