American Psychologist

ISSN: 0003-066X

2024, Vol. 79, No. 1, 137–149
https://doi.org/10.1037/amp0001168

Persons or Data Points? Ethics, Artificial Intelligence, and the Participatory Turn in Mental Health Research

Joshua August Skorburg[1], Kieran O'Doherty[2], and Phoebe Friesen[3]
[1] Department of Philosophy, University of Guelph
[2] Department of Psychology, University of Guelph
[3] Biomedical Ethics Unit, Department of Social Studies of Medicine, McGill University

This article identifies and examines a tension in mental health researchers' growing enthusiasm for the use of computational tools powered by advances in artificial intelligence and machine learning (AI/ML). Although there is increasing recognition of the value of participatory methods in science generally and in mental health research specifically, many AI/ML approaches, fueled by an ever-growing number of sensors collecting multimodal data, risk further distancing participants from research processes and rendering them as mere vectors or collections of data points. The imperatives of the "participatory turn" in mental health research may be at odds with the (often unquestioned) assumptions and data collection methods of AI/ML approaches. This article aims to show why this is a problem and how it might be addressed.

Public Significance Statement
Technologies powered by artificial intelligence (AI) and machine learning (ML) are transforming many aspects of society, including mental health research and treatment. We show why it is more important than ever to ensure that those most impacted by mental health research have a significant say in how these technologies are developed and deployed.

Keywords: artificial intelligence ethics, digital mental health, participatory research

This article identifies and examines a tension that arises in the context of mental health researchers' growing enthusiasm for the use of various computational tools powered by advances in artificial intelligence and machine learning (AI/ML). On the one hand, there is increasing recognition of the value of participatory methods in science generally and in mental health research specifically. But on the other hand, many AI/ML approaches, fueled by an ever-growing number of sensors collecting multimodal data, risk further distancing participants from research processes and rendering them as mere vectors or collections of data points. To put it bluntly, the imperatives of the "participatory turn" in mental health research may be at odds with the (often unquestioned) assumptions and data collection methods of AI/ML approaches. This article aims to show why this is a problem and how it might be addressed.

In The Participatory Turn section, we review some core commitments of the participatory turn in mental health research. Then, in the AI/ML Applications in Mental Health Research section, we review some recent developments in AI/ML methods used for mental health research. Next, in the Tensions Between AI/ML and Participatory Principles

Joshua August Skorburg https://orcid.org/0000-0002-3779-5076
Kieran O'Doherty https://orcid.org/0000-0002-9242-2061
Phoebe Friesen https://orcid.org/0000-0002-1529-916X
The authors have no conflicts of interest to declare.
Joshua August Skorburg played a lead role in conceptualization and an equal role in writing–original draft and writing–review and editing. Kieran O'Doherty played an equal role in writing–original draft and writing–review and editing. Phoebe Friesen played an equal role in writing–original draft and writing–review and editing.
Correspondence concerning this article should be addressed to Joshua August Skorburg, Department of Philosophy, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada. Email: skorburg@uoguelph.ca

section, we explore whether and to what extent these developments are compatible with the participatory principles described in The Participatory Turn section. We then conclude by describing several avenues through which participatory principles might be brought to bear in mental health research using AI/ML.

The Participatory Turn

The participatory turn in research can be situated within broader trends toward the democratization of knowledge, in which it is often argued that publics have a rightful place in the governance of science and technology (Burgess, 2014). The justification for the inclusion of broader publics in research rests on the observation that science relies heavily on public funding and public support and that it has broad ramifications for how societal decisions are made on important public policy issues. This shift toward democratizing knowledge can be seen in various initiatives such as participatory research, public and patient involvement and engagement, public deliberation, and citizen science, as well as requirements related to public or patient involvement in the practices of funders, editors, researchers, and research ethics boards.

In mental health research, calls for democratization, inclusion, and the rights of service users to participate in the construction of knowledge are nothing new. Such demands have long been part of efforts to resist the status quo in psychiatry, including the dominant focus on biomedical research to the neglect of structural factors, the frequent use of coercion and restraints, and the harmful and stigmatizing stereotypes often invoked. These efforts have been most prominent within the antipsychiatry movement, the recovery movement, and the consumer/survivor/ex-patient (c/s/x) movement, and also play a role in recent developments in survivor-led research and Mad studies (LeFrançois et al., 2013; Sweeney, 2016; Sweeney et al., 2009).[1] As such, it is worth considering the unique aspects of mental health research that underlie the participatory turn in this domain, before looking at how this turn may be in tension with developments in AI/ML. Below, we briefly outline both epistemic and ethical reasons for the inclusion of mental health service users, as well as members of the public who may become service users, in research relevant to them.[2]

Epistemic Considerations

Epistemic justifications underlying the participatory turn often emphasize positive, instrumental impacts that result from including those impacted by research in the research process itself. Studies related to this have proliferated in recent years, such that there are now several systematic reviews (Brett et al., 2014a, 2014b; Domecq et al., 2014).[3] While these reviews tend to differ in scope and focus, conclusions drawn often emphasize how participatory research approaches can improve the research process itself (e.g., identifying hurdles in advance, supporting recruitment, increasing dissemination opportunities), improve outcomes and data (e.g., as they may be more relevant to service users), and how stakeholders are impacted by involvement (e.g., through increased awareness, reflection, and understanding).

This focus on the impact of participation can be seen in mental health research as well and is often raised in response to worries that participatory research is less rigorous than traditional research. As Diana Rose has put it, there is a view that participatory research, "particularly in mental health, is biased, anecdotal and carried out by people who are over-involved" (Rose, 2014, p. 155). However, research related to the positive impacts of active engagement of patients and publics in research helps to counteract these worries, underlining an epistemic justification for involvement. For example, Gillard and colleagues report that involving mental health service users in their qualitative research project related to self-care led to the identification of novel themes and increased critical reflection of the research team (Gillard et al., 2010, 2012), while Simpson and House (2002) report that enlisting service users to conduct interviews "may have brought out negative opinions of services that would not otherwise have been obtained." Impacts related to increased enrollment, retention, improved methodologies, and the relevance of research have also been documented (Crocker et al., 2017, 2018; Domecq et al., 2014; Staley, 2009).

[1] While we use the terms "patients" and "service users" interchangeably in this article, we recognize that "patient" is a contested concept and that patiency is not a unified phenomenon. For example, some prefer to be referred to as clients, survivors, ex-patients, service users, and so forth. Following Tekin (2022, p. 4), what we have in mind with these terms is "the individual who is in a position of need due to the distress she is experiencing and who seeks help from a professional to address her condition." Similarly, it is worth noting how in many contexts the term "research participant" has come to replace earlier uses of terms such as "subjects" to denote humans who are the objects of scientific investigation. This use of the term "participant" thus positions human subjects of research as relatively passive, compared to the scientists who are conducting the research and who have the authority to make epistemic claims based on data collected about the subjects/participants. In contrast, when we speak of the "participatory turn" in mental health research, we are referring to calls for more active involvement of patients, service users, and broader publics in the production of knowledge. For research to be participatory, in this sense, means for subjects of research to also be engaged with more substantive aspects of the research process. The specific ways in which such involvement may occur vary and may include contributions to research design and analysis, decisions about funding research projects, and inclusion in the governance of research projects and infrastructure.

[2] While we refer separately to epistemic and ethical arguments below, such arguments are often overlapping and mutually reinforcing and cannot be cleanly untangled.

[3] Additional reviews have been led by funders focused on participatory research, like PCORI (the Patient-Centered Outcomes Research Institute) in the United States and INVOLVE in the United Kingdom (Forsythe et al., 2019; Staley, 2009).

These efforts to measure the impact of participatory research have grown significantly in recent years, particularly in jurisdictions where there are formal requirements for inclusion by institutions or funding calls (e.g., the National Institute for Health and Care Research in the United Kingdom and Canadian Institutes of Health Research in Canada).[4]

While there are various ways of theorizing which service users may be uniquely capable of contributing to mental health research, standpoint theory offers one helpful approach for understanding the basis of these claims related to expertise grounded in lived experience. The epistemic advantage identified by standpoint theory is also sometimes referred to as experiential expertise or experience by virtue of lived experience.

The basic premises of standpoint theory are that (a) those who have been marginalized often have unique epistemic potential related to their marginalization and (b) this potential can be activated through critical reflection, particularly related to power and knowledge. Standpoint theory asserts that marginalized individuals who have engaged in critical reflection have an epistemic advantage when it comes to knowledge projects relevant to their marginalization (Friesen & Goldstein, 2022; Figueroa et al., 2003).[5]

In mental health research, standpoint theory has been proposed as a justification for the importance of participatory research by several authors (Faulkner, 2017; Friesen & Goldstein, 2022; Rose, 2014, 2017). Their primary assertion is that lived experience of receiving a mental health diagnosis and/or engaging with the mental health system as a service user can provide one with an epistemic advantage, often involving the ability to identify problematic assumptions within a knowledge project, the ability to develop new hypotheses and theories or to operate with stronger objectivity in research[6] (Friesen & Goldstein, 2022; Figueroa et al., 2003).

Ethical Considerations

In contrast to arguments that focus on epistemic benefits of engaging patients in research, many of the ethical justifications for participatory research make the case that those being directly impacted by mental health research have a right to contribute to the guidance of such research, often captured by the motto: "Nothing about us without us." Arguments in this vein emphasize the importance of sharing power related to knowledge production with service users *because they deserve it* rather than because it is likely to benefit researchers or research. Possible foundations for this right may involve the unique stake service users have in research and how those receiving services are most likely to take on both risks and benefits within, and as a result of, research activities. Regardless of how it is understood, the right to be engaged in knowledge production, rather than a mere source of data, is well recognized; when surveyed, 90% of service users agreed that "service users have a fundamental right to actively participate as researchers in mental health research" (Patterson et al., 2014).

Justifications taking this shape are much more political in nature than those focused on epistemic reasons for involvement and are often tied to deep dissatisfaction and resistance to institutional mental health research and services. Mad studies, for instance, have been described as "the radical reclaiming of psychic spaces of resistance against the psychiatric domination of Mad people" (LeFrançois et al., 2013). Many have expressed worries about the likelihood of these more political, justice-oriented concerns being co-opted or swept under the rug in participatory projects led by academic centers or health care institutions, who may not wish to associate themselves with such overtly political or critical messages or aims (Pilgrim, 2005). This has led to increasing complaints about the "sanitization" of participatory research and practices related to selecting service user researchers as collaborators who are unlikely to disagree with the project as a whole (Faulkner, 2017; Russell et al., 2018).

Related justifications for the right of service users to contribute to mental health research focus on harms and abuse within the field. Indeed, the history of mental health research and practice is littered with examples of abuse and wrongdoing. Lists of treatments that were used in the past often sound like lists of torture techniques (e.g., lobotomies, bleeding, insulin comas, ice water baths, forced sterilizations). Research involving mental health patients was often abusive and unethical, from Nazi experiments to research conducted by the Imperial Japanese Army (López-Muñoz et al., 2007). Mental health diagnoses have often been used to silence or detain political dissidents or to further oppress marginalized groups, as in the famous cases of drapetomania, a diagnosis reserved for runaway slaves, or homosexuality, which was in the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) until 1973 (López-Muñoz et al., 2007; Spitzer, 1981; Willoughby, 2018).

Unfortunately, the harms prevalent within mental health research and practice are not only a thing of the past. Shackling and limited access to fundamental rights such as food, water, and daylight are far too commonly experienced by individuals deemed mentally ill in much of the Global South, while in the Global North, use of restraints (physical and chemical), seclusion, forced treatments, and police violence are common

[4] However, such a focus on instrumental justifications for inclusion often serves to obscure ethical reasons for involvement (Friesen et al., 2021). See Ethical Considerations section below.

[5] Well-known examples in other domains include female primatologists' ability to identify problematic assumptions guiding the field when they entered its ranks and Black women's contributions to sociology through the lens of intersectionality (Collins, 1986; Haraway, 1989).

[6] Strong objectivity is a form of objectivity espoused by Sandra Harding that involves explicitly acknowledging and reflecting on values that may inform one's methods and theories in a scientific project (Harding, 1992).

(Fatal Force, 2022; Frueh et al., 2005; Human Rights Watch, 2019, 2020). Such harmful and traumatizing forms of care are often justified through depictions of those diagnosed with mental health disorders as lacking capacity or rationality.[7]

These patterns of harm have led to deeply ingrained distrust and dissatisfaction with the mental health system and the research related to it (Geoffrey, Kloos & Ornelas, 2014). Ethical arguments for the participatory turn that foreground these harms highlight how, owing to hierarchies and power structures in mental health research, marginalized groups are rendered unable to speak for themselves on their own terms. In this way, participatory research can offer a corrective, balancing out the unfair distribution of power, in which one group of experts, those with academic training, are always in a position of speaking for others, despite the latter group's experiential expertise. As such, democratization of mental health research can be seen as protectionist, in that to prevent continued harm from occurring, it is important to have those most likely to experience such harms at the table. These arguments can also be read as reparative or restorative, in that shifting power to those who have been harmed can serve to repair or restore justice where it has been lost (Friesen et al., 2021).[8]

Participatory Principles

With these epistemic and ethical considerations in mind, we can extract two key commitments of the participatory turn as it relates to mental health research. This is not to say that these are the only two such commitments, but rather, that these are the most relevant for assessing the rising enthusiasm for the use of AI/ML in mental health research.

First, the foregoing highlights (a) ethical and epistemic commitments to the value and richness of lived experience in knowledge production about mental health. While it might seem obvious that the first-person experience of mental illness is relevant to mental health research, Tekin (2022) points out that the *DSM*—the primary classification system for mental health diagnosis—has never included service users as either "subjects" who generate research on mental illness nor as decision makers (e.g., members of *DSM* taskforce or working groups). Instead, "to the extent that they were part of the DSM's research into mental disorders, they have almost always been simply the 'objects' of investigation" (p. 3). In contrast, participatory approaches to mental health research recognize that there is a form of expertise generated from the lived experience of mental illness that is distinct from the more familiar clinical, diagnostic, and training-based forms of expertise. Such lived experience can be reflected not only in research design by including qualitative studies but also by including patients in decision-making roles and governance of mental health research.

Second, the foregoing highlights (b) ethical and epistemic commitments to stakeholders—whether individually or collectively—being able to speak for themselves. This follows directly from (a) because once the distinctive value of lived experience is recognized, the epistemic advantages gained from inhabiting particular standpoints can only be realized when occupants of those standpoints are able to describe their experience in their own terms. When exercised, this ability for participants to speak on their own terms is what delivers the benefits mentioned above, including uncovering novel themes in data, identifying barriers to implementation, increased study enrollment and retention, and so forth. For these benefits to be realized, mechanisms need to exist whereby stakeholders can articulate their perspectives, be heard by researchers, clinicians, and policy makers, and their perspectives be acted upon. Such spaces for dialogue do not occur naturally and must therefore be purposefully built into research processes and governance.

The AI/ML Applications in Mental Health Research

Having seen some of the epistemic and ethical justifications for participatory methods in mental health research, we turn now to a summary of some recent work using AI/ML in mental health research. The guiding question henceforth will be whether these new AI/ML approaches are likely to be compatible with the participatory justifications and principles just described.

It would be impossible to review all the recent developments in this rapidly evolving field (but for a helpful general methodological overview, see Garcia-Ceja et al., 2018), so our review is necessarily selective. Still, it aims to be representative of larger trends at the intersections of AI/ML and mental health, especially as these are relevant to the ethical and epistemic considerations described in The Participatory Turn section.

Language and Mental Health

Stanley Milgram said, "if the world were drained of every individual and we were left only with the messages that passed between them, we would still be in possession of the information needed to construct our discipline" (Milgram, 1977, p. 253, quoted in Dehghani & Boyd, 2022, p. xi). It is perhaps unsurprising, then, that one of the most active areas of research at the intersection of AI/ML and psychology generally, and

[7] Such depictions are increasingly being understood through the lens of epistemic injustice, in which individuals are deemed to lack credibility on the basis of prejudicial beliefs (Catala et al., 2021; Crichton et al., 2017).

[8] As noted above, some have objected that because participatory research is explicitly guided by values, it is therefore biased, less rigorous, and too subjective. Advocates of participatory research counter that merely pointing out that this research is guided by values is not a meaningful objection. Indeed, philosophers of science have argued that *all* research is guided by values (Douglas, 2009) and that reflecting on which values may be guiding a research program contributes to more objective science (Harding, 1992). Participatory researchers are often just being explicit about which values are guiding their work (i.e., justice, fairness, inclusion, etc.).

mental health specifically, involves various computational approaches to language analysis.

In a recent review of nearly 400 studies, Zhang et al. (2022) note that natural language processing (NLP) methods are used in many text corpora relevant to mental health, including social media posts and messages, transcripts of interviews, narrative writing, clinical notes, and electronic health records. NLP can be leveraged for information extraction, sentiment analysis, emotion detection, and mental health surveillance to automatically identify various mental health indicators relevant to providing support, early detection, prevention, diagnosis, and even treatment.

At a high level, these methods parse a variety of linguistic features (e.g., syntactic, semantic, and lexicon-based features, as well as affective/emotional features, statistical corpus-based features, along with temporal, social, and behavioral features; see Zhang et al., 2022, Table 2, for more details). More specifically, some influential and highly cited studies have used NLP methods to predict psychosis from transcribed interviews (Bedi et al., 2015), depression from Facebook posts (De Choudhury et al., 2013), and suicidality from Twitter posts (Coppersmith et al., 2018). Similar approaches have been used for identifying eating disorders among Reddit users (Yan et al., 2019), predicting self-injurious thoughts and behaviors on https://www.TeenHelp.org (Franz et al., 2020), detecting autism in spoken and written language (MacFarlane et al., 2022), along with a wide range of NLP approaches used to predict bipolar disorder (see Harvey et al., 2022, for a review).

While social media data mining approaches tend to garner the most attention, a recent review by Le Glaz et al. (2021) indicates that the most commonly used text corpora for mental health research are electronic health records, including both structured medical records and unstructured clinical notes. For this reason, we focus much of our analysis in the next section on these use cases. One early and influential study in this vein by Perlis et al. (2012) examined a population of over 127,000 patients diagnosed with major depressive disorder and found that using NLP on unstructured clinical notes better predicted current mood state than using structured billing data alone. Related approaches have used sentiment analysis and topic modeling on discharge notes to predict hospital readmission (McCoy et al., 2015; Rumshisky et al., 2016). More recent work in this area has refined and expanded these approaches for extracting mental health information from various text corpora (see, Le Glaz et al., 2021; Kariotis et al., 2022; Zhang et al., 2022; Zurynski et al., 2021, for relevant reviews).

Photos and Mental Health

Text generated by users on social media platforms (posts, comments, messages, etc.) has long provided a rich resource for mental health-related data mining, and much ink has already been spilled about the promises and perils of these approaches (see Chancellor & De Choudhury, 2020, for a critical review). We will turn to some of the criticisms below, but for now, it is worth noting that social media activity also generates other kinds of nonlinguistic data that are used in AI/ML applications.

In one striking demonstration, Reece and Danforth (2017) collected nearly 44,000 photographs from 166 Instagram users in an effort to identify markers of depression in posted photographs (71 of the 166 Instagram users reported a history of depression). The ML models used entirely computationally generated features, "including pixel analysis, face detection, and metadata parsing, which can be done at scale, without any additional human input" (Reece & Danforth, 2017, p. 3). More specifically, they extracted the total number of Instagram posts per day, the number of likes and comments on each photograph, the number of human faces in a photograph, as well as pixel-level averages of hue, saturation, and brightness. The best performing algorithm, a 100-tree random forest classifier, was able to reliably distinguish between posts made by depressed versus nondepressed users using only these computationally extracted features. Moreover, these signals of depression were shown to be detectable even before the date of users' first diagnosis of depression. Xu et al. (2020) used a similar approach, analyzing visual features of photos posted to Flickr with associated metadata (along with linguistic features) to distinguish between pre- and postonset of mental illness symptoms.

Typing, Movement, and Mental Health

In addition to linguistic and visual features, researchers are increasingly developing AI/ML tools to parse the streams of (meta)data generated from smart devices to identify potential markers of mental illness. The widespread adoption of various wearable technologies (smartwatches, smartrings, smartglasses, fitness trackers, etc.) has spawned a number of research efforts to use measures of heart rate variability (obtained via photoplethysmography) to detect and monitor stress and anxiety (see Hickey et al., 2021, for a review). Wearables also generate data about movement, physical activity, sleep quality, and circadian rhythms, such that researchers can build predictive models of depression risk (e.g., Rykov et al., 2021). And the integration of data from wearables with other forms of clinical and nonclinical data affords ML-driven personalized approaches to mental health treatment (e.g., Shah et al., 2021).

Related approaches leverage metadata generated by typing and movement to predict various mental health outcomes. For example, Cao et al. (2017) developed a deep-learning model to predict mood disruption and depression from mobile phone typing metadata. The researchers enrolled 40 subjects in an 8-week experiment (12 subjects were diagnosed with Bipolar I or II, and the rest were healthy controls). Subjects were given a mobile phone with a custom keyboard that collected typing

metadata for keypresses on alphanumeric characters, including duration of a keypress, time since last keypress, and distance from last key, as well as movement metadata from the phone's accelerometer. The goal was to compare the ability of different ML models to predict subjects' scores on the Hamilton Depression Rating Scale (HDRS) and the Young Mania Rating Scale (YMRS). The best performing model, which the authors dub "DeepMood," achieved over 90% accuracy. Impressively, the model delivers this level of performance when a subject provides around 400 typing "sessions" in the training phase (where a session is defined as beginning with a keypress that occurs after five or more seconds have elapsed since the last keypress and continuing until five or more seconds elapse between keypresses), where each "session" is typically less than a minute.

Last, Ware et al. (2020) built a family of ML models that use features extracted from WiFi infrastructure for large-scale depression detection. In a two-phase study, involving 182 college students (58 of which were diagnosed with depression), the researchers collected movement and location data from association logs that were captured at WiFi access points on a university campus. Many WiFi access points are associated with specific buildings and so data about visits to specific buildings, kinds of buildings (e.g., class, library, sports, entertainment), duration of stay, number of buildings visited, and so forth, can be gathered from the WiFi infrastructure without directly collecting any data from smartphones. The researchers used support vector machine models for each depressive symptom in the Patient Health Questionnaire-9 and Quick Inventory of Depressive Symptomology measures. They found that these movement and location features automatically extracted from WiFi metadata can accurately predict individual depressive symptoms (accuracy metrics were as high as 87%, but this varied by symptom).

A comprehensive review of AI/ML methods used in mental health research would take more space than we have here, and it would be outdated within a few months. Thus, our brief review here is meant to highlight some big-picture trends seen in the many modalities through which mental health-relevant data is generated, as well as the wide variety of AI/ML approaches used to potentially benefit clinicians and patients in mental health risk assessment, diagnosis, treatment, monitoring, and so forth. In the next section, we consider the relationship between these big-picture trends and the participatory turn described in The Participatory Turn section.

The Tensions Between AI/ML and Participatory Principles

When considering the ethical and epistemological implications of using AI/ML tools in mental health research, much of the scholarship has focused on issues like data privacy, informed consent, autonomy, efficacy, the differing ethical and legal standards in commercial versus academic contexts, and the (in)adequacy of the Belmont principles (Burr et al., 2020; Floridi & Cowls, 2022; Martinez-Martin & Kreitmair, 2018; Metcalf & Crawford, 2016; Morley et al., 2020; Skorburg & Friesen, 2021, 2022; Skorburg & Yam, 2022; Tekin, 2021; Torous & Roberts, 2017). Many of the ethical, legal, social, and epistemological implications described in this literature are relevant to the examples described in the AI/ML Applications in Mental Health Research section. But our aim here is to examine the extent to which the examples described in the AI/ML Applications in Mental Health Research section are compatible with the ability for stakeholders—whether individually or collectively—to speak for themselves and be heard on their own terms and also the recognition of the value and richness of lived experience. This is not to suggest that data privacy, informed consent, autonomy, efficacy, and regulatory considerations are somehow unimportant. Rather, we think that consideration of participatory approaches to research and research ethics can significantly advance these related discussions.

Speaking for Oneself

As we noted in the AI/ML Applications in Mental Health Research section, one of the most active areas of research at the intersection of AI/ML and mental health involves parsing various forms of linguistic and textual data. And as we noted in The Participatory Turn section, one of the core commitments of the participatory turn in mental health research is that participants ought to be able to speak for themselves, on their own terms. At first pass, it might seem like the increasing enthusiasm for NLP methods in mental health research would thus go hand in hand with the participatory turn. Careful consideration, however, reveals some worries about this compatibility.

Recall that the most commonly used text corpora for mental health research are electronic health records (EHR), including both structured medical records and unstructured clinical notes (Le Glaz et al., 2021). As Skorburg and Friesen (2022) have pointed out, text mining methods that make predictions and classifications based on EHR data often distance participants from the research process by relying on clinical notes and direct or indirect observations written by health care professionals. Almost by definition, the data encoded into the EHR in this way replace the patient's voice with those of clinicians. There are obvious concerns with various forms of bias being inscribed (e.g., van Ryn & Burke, 2000), but our main worry here has to do with the structural factors that can lead to the patient's voice being all but erased.

Consider a case where NLP researchers are tasked with predicting patient-level risk of rehospitalization based on clinical notes from EHRs. Very often, the AI/ML researchers building predictive models are not the ones entering clinical notes into the EHR. As such, the models are often (at least) two steps removed from the patient's experience. For example, if a

patient is describing a recent experience with social anxiety to a clinical psychologist, this experience is filtered and then summarized through the psychologist's notes entered into the EHR. In all likelihood, the psychologist has many such notes to write, and so writes them quickly, sometimes days after the initial session. The record for this particular patient might contain many other notes, albeit from different clinicians in different contexts. These can then be aggregated and combined with the records from different patients—collapsing across contexts, as it were—to identify linguistic features in clinical notes that are associated with higher rates of rehospitalization. By now, the patient's description of their experience of social anxiety, as summarized after the fact by their psychologist, is one of many vectors in a high-dimensional semantic space used to predict outcomes of interest for the hospital.

This is just a schematic example, but it illustrates how these kinds of methods could reveal surprising and important insights that can be used to improve patient care, optimize allocation of limited resources, and target interventions for those most in need. However, and this is our main contention, this often comes at the cost of substituting the patient's voice with the clinician's (and indeed, even the clinician's, with machine-predictable features). These are structural properties of EHR mining approaches in the sense that the massive amount of data required to generate useful predictions in turn requires a significant degree of input standardization, which often abstracts away from and erases the voice of the patient. In this way, the (largely unquestioned) background assumptions of much research in this vein serve to reify the distinction between expert researchers who have the epistemic authority to define experiences related to mental health, on the one hand, and those with mental illness as mere objects of study, on the other hand. Insofar as participatory approaches to mental health research, both challenge these assumptions about epistemic authority and also require patients to speak for themselves on their own terms, then many forms of EHR mining are going to be in tension with this participatory principle.

These worries about the incompatibility of EHR mining and participatory approaches are highlighted in a recent review by Kariotis et al. (2022), who note:

> An objective of EHRs is the standardization of health information to allow for health information exchange and data analytics. In comparison, mental health care involves the documentation of a large amount of narrative information, much of which resists standardization. An increasing focus on recovery models of mental health care that prioritize service user-defined measures and outcomes may create further tensions with standardized data collection. Concerns have also been raised about EHRs impeding clinicians' ability to understand a service user's entire story. … Future research and EHR design need to establish which standardized information is relevant for the mental health context and how best to present narrative information to capture service users' stories. (p. 14)

They also note that a significant limitation of any EHR mining approach to mental health is that clinicians frequently "water down" potentially sensitive information or keep such information in separate records. This is beneficial from the perspective of patient privacy. But these same practices may severely limit the development of accurate and clinically useful predictive models. Moreover, service users are not always able to access their own clinical notes, which leaves them unable to correct mistakes (see, e.g., Blease et al., 2021; Schwarz et al., 2021).

Most concerningly for our purposes, the authors note that in the studies included in their review, only 10% involved service users. Taken together, these considerations suggest that some of the most commonly used AI/ML methods for mental health research are incompatible with the commitments of the participatory turn and, hence, are unlikely to deliver the associated epistemic and ethical benefits described in The Participatory Turn section.[9]

First-Person Experience

As we saw in The Participatory Turn section, a second core commitment of the participatory turn in mental health research involves respecting the value and richness of first-person lived experience. Our guiding insight here is that many of the approaches described in the AI/ML Applications in Mental Health Research section require the use of stand-ins or proxies for mental health that abstract away from and subsequently devalue first-person experience. These dynamics can be subtle, but failing to appreciate them leads to the erasure of participants' voices, the solidification of problematic power dynamics, and an abstract, impersonal relationship between researchers and participants.

Consider the widespread practice of using AI/ML to predict scores on symptom scales like the HDRS. This is so common, and it is easy to forget that measures like the HDRS are stand-ins. Of course, there is nothing inherently wrong with the use of stand-ins. Probably, they are required by most forms of scientific inquiry. Serious ethical and epistemic concerns arise, however, when such proxies fail to be recognized as such and are treated as the very thing they are standing in for. This is a version of what William James famously called the "psychologist's fallacy."

Additionally, there is a long history of critique from service user communities related to the limitations of symptom scales in research. Such scales focus exclusively on symptoms, which are thought to represent the manifestation of illness,

[9] We hasten to add, however, that other NLP methods are explicit in their centering of patient voices. For example, Hart et al. (2020) used sentiment analysis on social media discussions related to psychotropic medications to determine what information patients were being exposed to online. These kinds of approaches and others like them are likely to be more compatible with participatory principles. Thus, there is nothing necessarily antiparticipatory about NLP. Still, in some of the most common use cases such as EHR mining, there is a persistent risk of structurally limiting the ability of patients to speak for themselves about their mental health or to be involved in the knowledge claims that are made about them and their condition.

but often fail to capture aspects of experience that may matter more to patients than the presence of symptoms; this might include anything from debilitating side effects, one's social support network, one's engagement in meaningful activities, or one's housing status. As such, research relying on symptom scales often implicitly dictates that what matters and is therefore worth measuring, is the presence of clinical symptoms rather than other features of service users' lives (Friesen, 2019). The push for the inclusion of broader scales focused on recovery and quality of life in research reflects this worry (Priebe, 2007).

Taken together, these considerations underscore how many of the approaches described in the AI/ML Applications in Mental Health Research section must rely on readily available and easy-to-collect stand-ins for mental illness, such as symptom scales (which themselves implicitly encode value judgments about what is important to measure and who has the authority to decide this). In turn, many AI/ML methods and workflows make it all too easy for researchers to fail to appreciate that a model is *actually* predicting these scores on symptom scales and not mental illness itself.

Consider the Instagram study reviewed above (Reece & Danforth, 2017), which automatically identified photogenic markers of depression. When considering these results through a participatory lens, one cannot help but notice that the automated, computational methods are processing stand-ins for stand-ins of the first-person experience of depression. We do not doubt that social media posts contain mental health-relevant information that can help to generate important insights. But it must also be said that social media posts represent *very* limited aspects of persons. They are often crude stand-ins for cognitive and affective states, subject to all manner of social desirability biases. What's more, in this study (and others like it) predictive power is derived from *stand-ins* of these stand-ins, such as pixel-level averages of hue, saturation, and brightness. Here also, we are (at least) two steps removed from the first-person experience of mental illness.[10] A similar analysis applies to the case of typing metadata (Cao et al., 2017). HDRS and YMRS scores are used as stand-ins for mental illness, which are predicted by features as impersonal as time between keypresses.

We worry that these methods systematically under-value first-person lived experience. The reliance on stand-ins for stand-ins of limited aspects of persons makes it all too easy for researchers to treat people like data points, rather than persons, precisely because the data used to train the models is so far removed from participants' first-person experience. And owing to the automated nature of the data collection and analysis, there is no room for the participants to express whether they feel adequately represented by the scales used, or whether there are other meaningful features the researchers ought to consider. Indeed, the automated, scalable nature of these methods is often touted as their primary benefit. For example, in the Ware et al. (2020) WiFi metadata study, the authors repeatedly highlight how the data are "collected passively without any efforts from the users" (p. 9). This sounds like the antithesis of the participatory approaches described in The Participatory Turn section. And finally, these examples exemplify the (often unacknowledged) gulf between algorithmic prediction and clinical intervention (Skorburg & Friesen, 2021). After all, nobody would suggest that decreasing time between keypresses or increasing color saturation on Instagram posts will reduce depressive symptoms.

It is worthwhile at this point to stop to consider an objection to our claims in this section: If these AI/ML approaches deliver accurate predictions, diagnoses, and so forth, then what is the harm in using them? If erasing participants' voices and abstracting away from their lived experience helps to generate better and more accurate predictions, the objection goes, that is a price worth paying. We contend that this line of thinking should be resisted in (at least) three ways.

First, it is at odds with the participatory turn (in science generally and mental health specifically) and its associated epistemic and ethical benefits, as we described them in The Participatory Turn section. Second, the objection expresses an instrumentalist attitude toward participants, which is incompatible with the foundational ethical principle of respect for persons. That is, it treats people as data points rather than persons. Third, as we saw above, most researchers do not bother to involve service users or end users in the research process. But when researchers have bothered to engage participants about, for example, algorithmic emotion detection on social media, participants' views of these methods were predominantly negative (Roemmich & Andalibi, 2021). Indeed, Reece and Danforth (2017, p. 10) note that "it is perhaps reflective of a current general skepticism towards social media research that, of the 509 individuals who began our survey, 221 (43%) refused to share their Instagram data, even after we provided numerous privacy guarantees." Of course, these claims do not necessarily impugn AI/ML research that does not employ social media mining. But the broader point about the need to meaningfully include the voices of service users in research still holds. This is especially clear when looking at trends in digital health more generally, where researchers have documented how feelings of distrust often arise when user data are collected surreptitiously; when the context of data collection is significantly different from the context of data analysis; and when third-party data brokers are involved (see, e.g., Johansson et al., 2021; Nwebonyi et al., 2022; Shah et al., 2021).[11]

[10] Unlike the majority of articles published in this field, and much to their credit, Reece and Danforth (2017) acknowledge this limitation in an endnote: "Occasionally, when reporting results we refer to 'observations' as 'participants,' e.g. 'depressed participants received fewer likes.' It would be more correct to use the phrase 'photographic data aggregated by participant-user-days' instead of 'participants.' We chose to sacrifice a degree of technical correctness for the sake of clarity" (p. 11).

[11] Thanks to the editors and reviewers for this point.

Two final points of clarification are needed. First, none of the above entails that we are somehow opposed to the use of AI/ML in mental health research. Second, we do not believe that participatory methods are a cure-all, nor that participatory methods are appropriate in all AI/ML research. Instead, we have aimed to highlight how some of the AI/ML methods and applications commonly used in mental health research tend to erase participants' voices and devalue their first-person experience. This is not a reason to eschew all AI/ML approaches in mental health research or to say these or similar worries do not arise in other forms of mental health research. But it is an invitation to more thoughtfully consider how to strike a balance between the benefits that AI/ML approaches offer on the one hand, with ethical and epistemic perils of treating people like data points, rather than persons, on the other hand. We also hope it will prompt consideration about which cases participatory methods may be most morally relevant (e.g., when the population being researched is especially vulnerable or has been exposed to considerable harm in the past).

In the concluding section, we briefly highlight some encouraging lines of research and offer some suggestions from successful participatory approaches in other domains.

Conclusion

Despite the tensions described above, there are several avenues through which participatory principles might be brought to bear in mental health research using AI/ML. First, there is a great potential for integrating participatory research methodologies, such as community-based participatory research (CBPR) or critical participatory action research (CPAR) into AI/ML research related to mental health (Fine et al., 2021; Holkup et al., 2004). CBPR has been described as "an approach to research that involves collective, reflective and systematic inquiry in which researchers and community stakeholders engage as equal partners in all steps of the research process with the goals of educating, improving practice or bringing about social change" (Tremblay et al., 2018). CPAR also seeks to engage those impacted by research in its production but is more political in nature, involving commitments to reframing the problem through critical theory, deep and broad participation, as well as action and accountability to social change and movements (Torre et al., 2012). These methods could be used to help counteract tendencies toward treating participants as data points that arise when AI/ML methods are utilized in mental health research. By involving service users and stakeholders in research projects from the start, so that those impacted by research have a say in the questions asked, the methods used, and the conclusions drawn, the risks of speaking on behalf of others can be minimized, and the benefits of participatory methods, be they ethical or epistemic, are more likely to be realized.

Second, there are growing trends related to participatory governance, in which those impacted by research are not merely involved in the research process, but are given authority to oversee and provide guidance on research protocols and proposals involving their community or population (del Campo et al., 2013; O'Doherty et al., 2011). These trends are seen most often in communities that are marginalized, overresearched, and have experienced harm within the context of research. Indigenous communities have made significant strides in this area, and communities of people who use drugs, disenfranchised neighborhoods, racialized communities, and those with rare conditions have followed suit, claiming a voice in research governance. Many mental health service users are also marginalized and have also experienced harm through research; as such, they may benefit from having a role in not only research involvement but in research oversight. This could involve drafting guidelines for mental health research involving AI/ML, participating in institutional review boards to review such research, or developing novel committees or working groups to provide guidance on how research could be structured or improved. One example of involving patients more directly in the design and execution of research is the RUDY (Rare U.K. Diseases of bone, joints, and blood vessels) study. The study utilizes a custom electronic platform through which patient partnership is built into the research design. Patients can upload disease and clinical history and experiences, which acknowledges that patients are best positioned to offer perspectives beyond just the clinical record. The platform also allows for dynamic consent for research participation. Most importantly, the governance structure of the study includes a patient forum consisting of 21 patients (Teare et al., 2017). Another example is the nonprofit organization Genetic Alliance. Because it is patient led, research supported by the organization is grounded in the experiences and interests of patients from the start. In addition, the governance structure of Genetic Alliance incorporates patients and community members in diverse roles. Broader patient engagement is enabled through an electronic platform.

Third, the burgeoning literature on algorithmic impact assessments promises to help bridge the gap between the data points collected in research and the lives that are likely to be impacted as a result. As Metcalf et al. (2021) note, "Algorithmic impact assessments (AIAs) are emerging governance practices for delineating accountability, rendering visible the harms caused by algorithmic systems, and ensuring practical steps are taken to ameliorate those harms" (p. 735). The devil is, of course, in the details, and Metcalf et al. (2021) highlight the difficulties with mapping the kinds of impacts within the scope of AIAs to the actual harms experienced by individuals or communities. Still, we are encouraged by some recent high-profile examples of AIAs being successfully deployed to limit potential harms in other domains such as hiring algorithms (e.g., O'Neil Risk Consulting and Algorithmic Auditing, 2020; Wilson et al., 2021). And indeed, many researchers are developing

AIA frameworks for health care (e.g., Lovelace Institute, 2022) that can be suitably adapted to the mental health context we are concerned with here. Relatedly, researchers and practitioners in the AI community are also increasingly recognizing the value of participatory approaches to AI system design (e.g., Birhane et al., 2022; Delgado et al., 2021; Donia & Shaw, 2021; Lee et al., 2019).

Finally, public deliberation has an important role in providing mechanisms for broader public input in policy guiding the use of AI/ML in mental health research. AI/ML is substantively changing the ways in which mental health research is conducted and how knowledge derived from this research is used to deliver mental health care. This has implications not only for current service users but all future users, their families, employers, and dependents. In short, AI/ML has profound implications for whole societies. Deliberative democratic principles hold that for societal decisions to be legitimate, they should be preceded by authentic deliberation among diverse members of the public (Dryzek & Niemeyer, 2010; Gutmann & Thompson, 2004). Public deliberation is an important mechanism that allows for the involvement of publics not only in sharing their perspectives and experiences but also in scrutinizing how trade-offs between competing values should be made and actively formulating recommendations for policy (O'Doherty, 2017). Public deliberation has now been used extensively and successfully for many controversial and emerging areas of science and technology (Burgess, 2014). The uptake of public deliberation as a form of practice by psychologists has been slow, though there is much promise in doing so (O'Doherty & Stroud, 2019). Laws and policies governing the design and implementation of AI/ML in mental health research should similarly be as cognisant as possible of ramifications for diverse publics and ensure congruence with broadly held values among (potentially) affected publics. To this end, some early promising efforts toward engaging publics in deliberative dialogue have been conducted, but further work in this regard will be important to ensure that public and patient voices have a place in guiding policy and conduct relating to mental health research and its applications.

References

Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., Ribeiro, S., Javitt, D. C., Copelli, M., & Corcoran, C. M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophrenia*, *1*(1), Article 15030. https://doi.org/10.1038/npjschz.2015.30

Birhane, A., Isaac, W., Prabhakaran, V., Díaz, M., Elish, M. C., Gabriel, I., & Mohamed, S. (2022). Power to the people? Opportunities and challenges for participatory AI. *Equity and access in algorithms, mechanisms, and optimization* (pp. 1–8). Association for Computing Machinery.

Blease, C., Salmi, L., Rexhepi, H., Hägglund, M., & DesRoches, C. M. (2021). Patients, clinicians and open notes: Information blocking as a case of epistemic injustice. *Journal of Medical Ethics*, *48*(10), 785–793. https://doi.org/10.1136/medethics-2021-107275

Brett, J., Staniszewska, S., Mockford, C., Herron-Marx, S., Hughes, J., Tysall, C., & Suleman, R. (2014a). A systematic review of the impact of patient and public involvement on service users, researchers and communities. *Patient*, *7*(4), 387–395. https://doi.org/10.1007/s40271-014-0065-0

Brett, J., Staniszewska, S., Mockford, C., Herron-Marx, S., Hughes, J., Tysall, C., & Suleman, R. (2014b). Mapping the impact of patient and public involvement on health and social care research: A systematic review. *Health Expectations*, *17*(5), 637–650. https://doi.org/10.1111/j.1369-7625.2012.00795.x

Burgess, M. M. (2014). From 'trust us' to participatory governance: Deliberative publics and science policy. *Public Understanding of Science*, *23*(1), 48–52. https://doi.org/10.1177/0963662512472160

Burr, C., Morley, J., Taddeo, M., & Floridi, L. (2020). Digital psychiatry: Risks and opportunities for public health and wellbeing. *IEEE Transactions on Technology and Society*, *1*(1), 21–33. https://doi.org/10.1109/TTS.2020.2977059

Cao, B., Zheng, L., Zhang, C., Yu, P. S., Piscitello, A., Zulueta, J., Ajilore, O., Ryan, K., & Leow, A. D. (2017, August). *Deepmood: Modeling mobile phone typing dynamics for mood detection* [Conference session]. Proceedings of the 23rd ACM International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada.

Catala, A., Faucher, L., & Poirier, P. (2021). Autism, epistemic injustice, and epistemic disablement: A relational account of epistemic agency. *Synthese*, *199*(3), 9013–9039. https://doi.org/10.1007/s11229-021-03192-7

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. *npj Digital Medicine*, *3*(1), Article 43. https://doi.org/10.1038/s41746-020-0233-7

Collins, P. H. (1986). Learning from the outsider within: The sociological significance of Black feminist thought. *Social Problems*, *33*(6), s14–s32. https://doi.org/10.2307/800672

Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural language processing of social media as screening for suicide risk. *Biomedical Informatics Insights*, *10*. https://doi.org/10.1177/1178222618792860

Crichton, P., Carel, H., & Kidd, I. J. (2017). Epistemic injustice in psychiatry. *BJPsych Bulletin*, *41*(2), 65–70. https://doi.org/10.1192/pb.bp.115.050682

Crocker, J. C., Boylan, A. M., Bostock, J., & Locock, L. (2017). Is it worth it? Patient and public views on the impact of their involvement in health research and its assessment: A UK-based qualitative interview study. *Health Expectations*, *20*(3), 519–528. https://doi.org/10.1111/hex.12479

Crocker, J. C., Ricci-Cabello, I., Parker, A., Hirst, J. A., Chant, A., Petit-Zeman, S., Evans, D., & Rees, S. (2018). Impact of patient and public involvement on enrolment and retention in clinical trials: Systematic review and meta-analysis. *The BMJ*, *363*, Article k4738. https://doi.org/10.1136/bmj.k4738

De Choudhury, M., Counts, S., & Horvitz, E. (2013, May). *Social media as a measurement tool of depression in populations* [Conference session]. Proceedings of the 5th Annual ACM Web Science Conference, Paris, France. https://dl.acm.org/doi/10.1145/2464464.2464480

Dehghani, M., & Boyd, R. L. (2022). Preface. In M. Dehghani & R. L. Boyd (Eds.), *Handbook of language analysis in psychology* (pp. xi–xii). Guilford Press.

del Campo, F. M., Casado, J., Spencer, P., & Strelnick, H. (2013). The development of the Bronx Community Research Review Board: A pilot feasibility project for a model of community consultation. *Progress in Community Health Partnerships*, *7*(3), 341–352. https://doi.org/10.1353/cpr.2013.0037

Delgado, F., Yang, S., Madaio, M., & Yang, Q. (2021). *Stakeholder participation in AI: Beyond "add diverse stakeholders and stir"*. arXiv preprint. https://doi.org/10.48550/arXiv.2111.01122

Domecq, J. P., Prutsky, G., Elraiyah, T., Wang, Z., Nabhan, M., Shippee, N., Brito, J. P., Boehmer, K., Hasan, R., Firwana, B., Erwin, P., Eton, D., Sloan, J., Montori, V., Asi, N., Dabrh, A. M., & Murad, M. H. (2014).

Patient engagement in research: A systematic review. *BMC Health Services Research*, *14*(1), Article 89. https://doi.org/10.1186/1472-6963-14-89

Donia, J., & Shaw, J. A. (2021). Co-design and ethical artificial intelligence for health: An agenda for critical research and practice. *Big Data & Society*, *8*(2). https://doi.org/10.1177/20539517211065248

Douglas, H. E. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Press. https://doi.org/10.2307/j.ctt6wrc78

Dryzek, J., & Niemeyer, S. (2010). *Foundations and frontiers of deliberative governance*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199562947.001.0001

Fatal Force. (2022). 1,066 people have been shot and killed by police in the past 12 months. *The Washington Post*. https://www.washingtonpost.com/graphics/investigations/police-shootings-database/

Faulkner, A. (2017). Survivor research and mad studies: The role and value of experiential knowledge in mental health research. *Disability & Society*, *32*(4), 500–520. https://doi.org/10.1080/09687599.2017.1302320

Figueroa, R., Harding, S., & Harding, S. G. (Eds.). (2003). *Science and other cultures: Issues in philosophies of science and technology*. Routledge.

Fine, M., Torre, M. E., Oswald, A. G., & Avory, S. (2021). Critical participatory action research: Methods and praxis for intersectional knowledge production. *Journal of Counseling Psychology*, *68*(3), 344–356. https://doi.org/10.1037/cou0000445

Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design* (pp. 535–545). Wiley.

Forsythe, L. P., Carman, K. L., Szydlowski, V., Fayish, L., Davidson, L., Hickam, D. H., Hall, C., Bhat, G., Neu, D., Stewart, L., Jalowsky, M., Aronson, N., & Anyanwu, C. U. (2019). Patient engagement in research: Early findings from the patient-centered outcomes research institute. *Health Affairs*, *38*(3), 359–367. https://doi.org/10.1377/hlthaff.2018.05067

Franz, P. J., Nook, E. C., Mair, P., & Nock, M. K. (2020). Using topic modeling to detect and describe self-injurious and related content on a large-scale digital platform. *Suicide & Life-Threatening Behavior*, *50*(1), 5–18. https://doi.org/10.1111/sltb.12569

Friesen, P. (2019). Expanding outcome measures in schizophrenia research: Does the research domain criteria pose a threat? *Philosophy, Psychiatry, & Psychology*, *26*(3), 243–260. https://doi.org/10.1353/ppp.2019.0039

Friesen, P., & Goldstein, J. (2022). Standpoint theory and the psy sciences: Can marginalization and critical engagement lead to an epistemic advantage? *Hypatia*, *37*(4), 659–687. https://doi.org/10.1017/hyp.2022.58

Friesen, P., Lignou, S., Sheehan, M., & Singh, I. (2021). Measuring the impact of participatory research in psychiatry: How the search for epistemic justifications obscures ethical considerations. *Health Expectations*, *24*(Suppl. 1), 54–61. https://doi.org/10.1111/hex.12988

Frueh, B. C., Knapp, R. G., Cusack, K. J., Grubaugh, A. L., Sauvageot, J. A., Cousins, V. C., Yim, E., Robins, C. S., Monnier, J., & Hiers, T. G. (2005). Patients' reports of traumatic or harmful experiences within the psychiatric setting. *Psychiatric Services*, *56*(9), 1123–1133. https://doi.org/10.1176/appi.ps.56.9.1123

Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K. J., & Tørresen, J. (2018). Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing*, *51*, 1–26. https://doi.org/10.1016/j.pmcj.2018.09.003

Geoffrey, N., Kloos, B., & Ornelas, J. (Eds.). (2014). *Community psychology and community mental health: Towards transformative change*. Oxford Academic. https://doi.org/10.1093/acprof:oso/9780199362424.001.0001

Gillard, S., Borschmann, R., Turner, K., Goodrich-Purnell, N., Lovell, K., & Chambers, M. (2010). 'What difference does it make?' Finding evidence of the impact of mental health service user researchers on research into the experiences of detained psychiatric patients. *Health Expectations*, *13*(2), 185–194. https://doi.org/10.1111/j.1369-7625.2010.00596.x

Gillard, S., Simons, L., Turner, K., Lucock, M., & Edwards, C. (2012). Patient and public involvement in the coproduction of knowledge: Reflection on the analysis of qualitative data in a mental health study. *Qualitative Health Research*, *22*(8), 1126–1137. https://doi.org/10.1177/1049732312448541

Gutmann, A., & Thompson, D. (2004). *Why deliberative democracy?* Princeton University Press. https://doi.org/10.1515/9781400826339

Haraway, D. (1989). *Primate visions: Gender, race, and nature in the world of modern science*. Psychology Press.

Harding, S. (1992). Rethinking standpoint epistemology: What is "strong objectivity"? *The Centennial Review*, *36*(3), 437–470.

Hart, K. L., Perlis, R. H., & McCoy, T. H., Jr. (2020). What do patients learn about psychotropic medications on the web? A natural language processing study. *Journal of Affective Disorders*, *260*, 366–371. https://doi.org/10.1016/j.jad.2019.09.043

Harvey, D., Lobban, F., Rayson, P., Warner, A., & Jones, S. (2022). Natural language processing methods and bipolar disorder: Scoping review. *JMIR Mental Health*, *9*(4), Article e35928. https://doi.org/10.2196/35928

Hickey, B. A., Chalmers, T., Newton, P., Lin, C. T., Sibbritt, D., McLachlan, C. S., Clifton-Bligh, R., Morley, J., & Lal, S. (2021). Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review. *Sensors*, *21*(10), Article 3461. https://doi.org/10.3390/s21103461

Holkup, P. A., Tripp-Reimer, T., Salois, E. M., & Weinert, C. (2004). Community-Based participatory research: An approach to intervention research with a Native American community. *ANS. Advances in nursing science*, *27*(3), 162–175. https://doi.org/10.1097/00012272-200407000-00002

Human Rights Watch. (2019). *"Fading away": How aged care facilities in Australia chemically restrain older people with dementia*. https://www.hrw.org/report/2019/10/15/fading-away/how-aged-care-facilities-australia-chemically-restrain-older-people

Human Rights Watch. (2020). *Living in chains: Shackling of people with psychosocial disabilities worldwide*.

Johansson, J. V., Bentzen, H. B., Shah, N., Haraldsdóttir, E., Jónsdóttir, G. A., Kaye, J., Mascalzoni, D., & Veldwijk, J. (2021). Preferences of the public for sharing health data: Discrete choice experiment. *JMIR Medical Informatics*, *9*(7), Article e29614. https://doi.org/10.2196/29614

Kariotis, T. C., Prictor, M., Chang, S., & Gray, K. (2022). Impact of electronic health records on information practices in mental health contexts: Scoping review. *Journal of Medical Internet Research*, *24*(5), Article e30405. https://doi.org/10.2196/30405

Lee, M. K., Kusbit, D., Kahng, A., Kim, J. T., Yuan, X., Chan, A., See, D., Noothigattu, R., Lee, S., Psomas, A., & Procaccia, A. D. (2019). *WeBuildAI: Participatory framework for algorithmic governance* [Conference session]. Proceedings of the ACM on Human–Computer Interaction. https://dl.acm.org/doi/10.1145/3359283

LeFrançois, B. A., Menzies, R., & Reaume, G. (Eds.). (2013). *Mad matters: A critical reader in Canadian mad studies*. Canadian Scholars' Press.

Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., DeVylder, J., Walter, M., Berrouiguet, S., & Lemey, C. (2021). Machine learning and natural language processing in mental health: Systematic review. *Journal of Medical Internet Research*, *23*(5), Article e15708. https://doi.org/10.2196/15708

López-Muñoz, F., Alamo, C., Dudley, M., Rubio, G., García-García, P., Molina, J. D., & Okasha, A. (2007). Psychiatry and political-institutional abuse from the historical perspective: The ethical lessons of the Nuremberg Trial on their 60th anniversary. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *31*(4), 791–806. https://doi.org/10.1016/j.pnpbp.2006.12.007

Lovelace Institute. (2022, February). *Algorithmic impact assessment: A case study in healthcare*. https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/

MacFarlane, H., Salem, A. C., Chen, L., Asgari, M., & Fombonne, E. (2022). Combining voice and language features improves automated autism detection. *Autism Research*, *15*(7), 1288–1300. https://doi.org/10.1002/aur.2733

Martinez-Martin, N., & Kreitmair, K. (2018). Ethical issues for direct-to-consumer digital psychotherapy apps: Addressing accountability, data protection, and consent. *JMIR Mental Health*, *5*(2), Article e32. https://doi.org/10.2196/mental.9423

McCoy, T. H., Castro, V. M., Cagan, A., Roberson, A. M., Kohane, I. S., & Perlis, R. H. (2015). Sentiment measured in hospital discharge notes is associated with readmission and mortality risk: An electronic health record study. *PLOS ONE*, *10*(8), Article e0136341. https://doi.org/10.1371/journal.pone.0136341

Metcalf, J., & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951716650211

Metcalf, J., Moss, E., Watkins, E. A., Singh, R., & Elish, M. C. (2021). *Algorithmic impact assessments and accountability: The co-construction of impacts* [Conference session]. FAccT '21, March 3–10, 2021, Virtual Event, Canada. https://dl.acm.org/doi/pdf/10.1145/3442188.3445935

Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine*, *260*, Article 113172. https://doi.org/10.1016/j.socscimed.2020.113172

Nwebonyi, N., Silva, S., & de Freitas, C. (2022). Public views about involvement in decision-making on health data sharing, access, use and reuse: The importance of trust in science and other institutions. *Frontiers in Public Health*, *10*, Article 852971. https://doi.org/10.3389/fpubh.2022.852971

O'Doherty, K. C. (2017). Deliberative public opinion: Development of a social construct. *History of the Human Sciences*, *30*(4), 124–145. https://doi.org/10.1177/0952695117722718

O'Doherty, K. C., Burgess, M. M., Edwards, K., Gallagher, R. P., Hawkins, A. K., Kaye, J., McCaffrey, V., & Winickoff, D. E. (2011). From consent to institutions: Designing adaptive governance for genomic biobanks. *Social Science & Medicine*, *73*(3), 367–374. https://doi.org/10.1016/j.socscimed.2011.05.046

O'Doherty, K. C., & Stroud, K. (2019). Public deliberation and social psychology: Integrating theories of participation with social psychological research and practice. In K. O'Doherty & D. Hodgetts (Eds.), *The SAGE handbook of applied social psychology* (pp. 419–444). SAGE Publications.

O'Neil Risk Consulting and Algorithmic Auditing. (2020). *ORCAA's algorithmic audit of HireVue—Description of algorithmic audit: Pre-built assessments*. https://www.hirevue.com/resources/orcaa-report

Patterson, S., Trite, J., & Weaver, T. (2014). Activity and views of service users involved in mental health research: UK survey. *The British Journal of Psychiatry*, *205*(1), 68–75. https://doi.org/10.1192/bjp.bp.113.128637

Perlis, R. H., Iosifescu, D. V., Castro, V. M., Murphy, S. N., Gainer, V. S., Minnier, J., Cai, T., Goryachev, S., Zeng, Q., Gallagher, P. J., Fava, M., Weilburg, J. B., Churchill, S. E., Kohane, I. S., & Smoller, J. W. (2012). Using electronic medical records to enable large-scale studies in psychiatry: Treatment resistant depression as a model. *Psychological Medicine*, *42*(1), 41–50. https://doi.org/10.1017/S0033291711000997

Pilgrim, D. (2005). Protest and co-option—The voice of mental health service users. In A. Bell & P. Lindley (Eds.), *Beyond the water towers: The unfinished revolution in mental health services* (Vol. *2005*, pp. 41–42). The Sainsbury Centre for Mental Health

Priebe, S. (2007). Social outcomes in schizophrenia. *The British Journal of Psychiatry*, *191*(S50), S15–S20. https://doi.org/10.1192/bjp.191.50.s15

Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, *6*(1), Article 15. https://doi.org/10.1140/epjds/s13688-017-0110-z

Roemmich, K., & Andalibi, N. (2021). *Data subjects' conceptualizations of and attitudes toward automatic emotion recognition-enabled wellbeing interventions on social media* [Conference session]. Proceedings of the ACM on Human–Computer Interaction. https://dl.acm.org/doi/abs/10.1145/3476049

Rose, D. (2014). Patient and public involvement in health research: Ethical imperative and/or radical challenge? *Journal of Health Psychology*, *19*(1), 149–158. https://doi.org/10.1177/1359105313500249

Rose, D. (2017). Service user/survivor-led research in mental health: Epistemological possibilities. *Disability & Society*, *32*(6), 773–789. https://doi.org/10.1080/09687599.2017.1320270

Rumshisky, A., Ghassemi, M., Naumann, T., Szolovits, P., Castro, V. M., McCoy, T. H., & Perlis, R. H. (2016). Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Translational Psychiatry*, *6*(10), e921. https://doi.org/10.1038/tp.2015.182

Russell, G., Starr, S., Elphick, C., Rodogno, R., & Singh, I. (2018). Selective patient and public involvement: The promise and perils of pharmaceutical intervention for autism. *Health Expectations*, *21*(2), 466–473. https://doi.org/10.1111/hex.12637

Rykov, Y., Thach, T. Q., Bojic, I., Christopoulos, G., & Car, J. (2021). Digital biomarkers for depression screening with wearable devices: Cross-sectional study with machine learning modeling. *JMIR mHealth and uHealth*, *9*(10), Article e24872. https://doi.org/10.2196/24872

Schwarz, J., Bärkås, A., Blease, C., Collins, L., Hägglund, M., Markham, S., & Hochwarter, S. (2021). Sharing clinical notes and electronic health records with people affected by mental health conditions: Scoping review. *JMIR Mental Health*, *8*(12), Article e34170. https://doi.org/10.2196/34170

Shah, R. V., Grennan, G., Zafar-Khan, M., Alim, F., Dey, S., Ramanathan, D., & Mishra, J. (2021). Personalized machine learning of depressed mood using wearables. *Translational Psychiatry*, *11*(1), Article 338. https://doi.org/10.1038/s41398-021-01445-0

Simpson, E. L., & House, A. O. (2002). Involving users in the delivery and evaluation of mental health services: Systematic review. *The BMJ*, *325*(7375), Article 1265. https://doi.org/10.1136/bmj.325.7375.1265

Skorburg, J. A., & Friesen, P. (2021). Mind the gaps: Ethical and epistemic issues in the digital mental health response to COVID-19. *The Hastings Center Report*, *51*(6), 23–26. https://doi.org/10.1002/hast.1292

Skorburg, J. A., & Friesen, P. (2022). Ethical issues in text mining for mental health. In M. Dehghani & R. Boyd (Eds.), *The atlas of language analysis in psychology* (pp. 531–550). Guilford Press.

Skorburg, J. A., & Yam, J. (2022). Is there an app for that?: Ethical issues in the digital mental health response to COVID-19. *AJOB Neuroscience*, *13*(3), 177–190. https://doi.org/10.1080/21507740.2021.1918284

Spitzer, R. L. (1981). The diagnostic status of homosexuality in DSM-III: A reformulation of the issues. *The American Journal of Psychiatry*, *138*(2), 210–215. https://doi.org/10.1176/ajp.138.2.210

Staley, K. (2009). *Exploring impact: Public involvement in NHS, public health and social care research*. Eastleigh.

Sweeney, A. (2016). Why mad studies needs survivor research and survivor research needs mad studies. *Intersectionalities: A Global Journal of Social Work Analysis, Research, Polity, and Practice*, *5*(3), 36–61.

Sweeney, A., Beresford, P., Faulkner, A., Nettle, M., & Rose, D. (2009). *This is survivor research*. PCCS Books.

Teare, H. J. A., Hogg, J., Kaye, J., Luqmani, R., Rush, E., Turner, A., Watts, L., Williams, M., & Javaid, M. K. (2017). The RUDY study: Using digital technologies to enable a research partnership. *European Journal of Human Genetics*, *25*(7), 816–822. https://doi.org/10.1038/ejhg.2017.57

Tekin, Ş. (2021). Is big data the new stethoscope? Perils of digital phenotyping to address mental illness. *Philosophy & Technology*, *34*(3), 447–461. https://doi.org/10.1007/s13347-020-00395-7

Tekin, Ş. (2022). Participatory interactive objectivity in psychiatry. *Philosophy of Science*, *89*(5), 1166–1175. https://doi.org/10.1017/psa.2022.47

Torous, J., & Roberts, L. W. (2017). Needed innovation in digital health and smartphone applications for mental health: Transparency and trust. *JAMA Psychiatry*, *74*(5), 437–438. https://doi.org/10.1001/jamapsychiatry.2017.0262

Torre, M. E., Fine, M., Stoudt, B. G., & Fox, M. (2012). Critical participatory action research as public science. In H. Cooper, P. M. Camic, D. L. Long,

A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology* (pp. 171–184). American Psychological Association. https://doi.org/10.1037/13620-011

Tremblay, M. C., Martin, D. H., McComber, A. M., McGregor, A., & Macaulay, A. C. (2018). Understanding community-based participatory research through a social movement framework: A case study of the Kahnawake Schools Diabetes Prevention Project. *BMC Public Health*, *18*(1), Article 487. https://doi.org/10.1186/s12889-018-5412-y

van Ryn, M., & Burke, J. (2000). The effect of patient race and socio-economic status on physicians' perceptions of patients. *Social Science & Medicine*, *50*(6), 813–828. https://doi.org/10.1016/S0277-9536(99)00338-X

Ware, S., Yue, C., Morillo, R., Lu, J., Shang, C., Bi, J., Kamath, J., Russell, A., Bamis, A., & Wang, B. (2020). Predicting depressive symptoms using smartphone data. *Smart Health*, *15*, Article 100093. https://doi.org/10.1016/j.smhl.2019.100093

Willoughby, C. D. (2018). Running away from Drapetomania: Samuel A. Cartwright, medicine, and race in the Antebellum South. *The Journal of Southern History*, *84*(3), 579–614. https://doi.org/10.1353/soh.2018.0164

Wilson, C., Ghosh, A., Jiang, S., Mislove, A., Baker, L., Szary, J., Trindel, K., & Polli, F. (2021, March). *Building and auditing fair algorithms: A case study in candidate screening* [Conference session]. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada.

Xu, Z., Pérez-Rosas, V., & Mihalcea, R. (2020, May). *Inferring social media users' mental health status from multimodal information* [Conference session]. Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France. https://aclanthology.org/2020.lrec-1.772/

Yan, H., Fitzsimmons-Craft, E. E., Goodman, M., Krauss, M., Das, S., & Cavazos-Rehg, P. (2019). Automatic detection of eating disorder-related social media posts that could benefit from a mental health intervention. *International Journal of Eating Disorders*, *52*(10), 1150–1156. https://doi.org/10.1002/eat.23148

Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: A narrative review. *npj Digital Medicine*, *5*(1), Article 46. https://doi.org/10.1038/s41746-022-00589-7

Zurynski, Y., Ellis, L. A., Tong, H. L., Laranjo, L., Clay-Williams, R., Testa, L., Meulenbroeks, I., Turton, C., & Sara, G. (2021). Implementation of electronic medical records in mental health settings: Scoping review. *JMIR Mental Health*, *8*(9), Article e30564. https://doi.org/10.2196/30564

Received August 16, 2022
Revision received January 18, 2023
Accepted April 3, 2023 ■