



# Detecting your depression with your smartphone? – An ethical analysis of epistemic injustice in passive self-tracking apps

Mirjam Faissner<sup>1,2</sup> · Eva Kuhn<sup>3</sup> · Regina Müller<sup>4</sup> · Sebastian Laacke<sup>5</sup>

© The Author(s) 2024

## Abstract

Smartphone apps might offer a low-threshold approach to the detection of mental health conditions, such as depression. Based on the gathering of ‘passive data,’ some apps generate a user’s ‘digital phenotype,’ compare it to those of users with clinically confirmed depression and issue a warning if a depressive episode is likely. These apps can, thus, serve as epistemic tools for affected users. From an ethical perspective, it is crucial to consider epistemic injustice to promote socially responsible innovations within digital mental healthcare. In cases of epistemic injustice, people are wronged specifically as epistemic agents, i.e., agents of the production and distribution of knowledge. We suggest that epistemic agency relies on different resource- and uptake-related preconditions which can be impacted by the functionality of passive self-tracking apps. We consider how this can lead to different forms of epistemic injustice (testimonial, hermeneutical, and contributory injustice) and analyze the influence of the apps’ use on epistemic practices on an individual level, in mental healthcare settings, and on the structural level.

**Keywords** Digital phenotyping · e-mental health · Medical ethics · Epistemic (in)justice · Mobile health · Depression

## Introduction

One major concern in mental healthcare is the treatment and diagnosis gap. The rate for the minimally adequate treatment of major depressive disorder is as low as 23% in high income countries (Moitra et al., 2022). Many people with depression remain undiagnosed and, therefore, do not

receive recognition or treatment. In addition to a lack of mental health literacy, access barriers impede contact with mental healthcare services. Among these constraints are direct or indirect costs (e.g., for public transport), restricted mobility (e.g., due to long distances to the nearest psychiatrist/psychotherapist) and long waiting times (Köhnen et al., 2019). Digital mental health, and specifically smartphone applications (apps), have been suggested to offer a low-threshold approach to diagnosis and psychotherapy, thus, remedying the treatment and diagnosis gap (Fiske et al., 2019; Mayer et al., 2019), and promote justice in the healthcare system. The apps “may address these constraints [of mobility, cost, and motivation] and provide better screening for depression” (Al Hanai et al., 2018, p. 1716).

This article provides an ethical analysis of questions of epistemic injustice that arise in the context of one emerging technology: so-called ‘digital phenotyping’ apps. Epistemic injustice describes a “wrong done to someone specifically in their capacity as a knower” (Fricker, 2007, p. 1), and examines how people are wronged specifically as epistemic agents, i.e., agents of the production and distribution of knowledge (Dotson, 2014). Digital phenotyping apps are typically based on an artificial multimodal neural network

✉ Mirjam Faissner  
mirjam.faissner@charite.de

Eva Kuhn  
eva.kuhn@charite.de

<sup>1</sup> Institute of the History of Medicine and Ethics in Medicine, Charité – Universitätsmedizin Berlin, Thielallee 71, 14195 Berlin, Germany

<sup>2</sup> Department of Psychiatry, Psychotherapy and Preventive Medicine, Ruhr-University Bochum, Bochum, Germany

<sup>3</sup> Department of Psychiatry and Neurosciences, Charité – Universitätsmedizin Berlin, Berlin, Germany

<sup>4</sup> Institute of Philosophy, University of Bremen, Bremen, Germany

<sup>5</sup> Institute of Ethics and the History of Medicine, University Medicine Greifswald, Greifswald, Germany

model,<sup>1</sup> and analyze text and/or audio sequences harvested from social media, video data, data from global positioning systems, or any other interaction with one's mobile device (Al Hanai et al., 2018; Su et al., 2020). The data is analyzed for voice tone, facial expressions, content of text and speech, as well as other similar characteristics, and translated into clinically relevant 'digital biomarkers' (Jain et al., 2015; Torous et al., 2021). The latter are understood as "objective, quantifiable, physiological, and behavioral measures that are collected by means of digital devices that are portable, wearable, implantable, or ingestible" (Babrak et al., 2019, p. 93). The data is collected at any time the smartphone is not completely switched off (Torous et al., 2021). All the traces users leave behind when 'moving' around in the digital space and interacting with their device result in a 'digital phenotype' of the individual user (Spinazze et al., 2019). The digital phenotype can be understood as "one's digital footprint" (Jain et al., 2015, p. 463), which is expected to reveal different kinds of information about the app user, including indicators of their mental states (Leaning et al., 2024).

This technology is currently being developed for apps to predict the onset (or recurrence) of depressive episodes,<sup>2</sup> based on a user's digital phenotype. In this model, the app's algorithm compares a user's current digital phenotype to those of users with clinically confirmed depression and issues a warning if a depressive episode is likely. Many existing mental health self-tracking apps already on the market, such as Mindstrong or Behaviorance, use digital phenotyping as one among other functionalities, such as data collected based on user-app interaction, for example, by filling in questionnaires or keeping a mood diary (Leaning et al., 2024; Polhemus et al., 2022). The main feature that distinguishes the apps under scrutiny in this article is that they *exclusively* use digital phenotyping to assess a user's mental state. Such apps are currently being trialed

in clinical studies: A recent systematic review on the use of digital phenotyping in the context of depression identified 24 studies that exclusively used passive data to predict the onset of depressive episodes in users with a formal diagnosis of depression (Leaning et al., 2024). In these studies, the apps' output based on digital phenotyping is correlated with clinical measures assessing depressive symptoms, suggesting that the possibility of replacing clinical assessments with digital phenotyping is investigated. Digital phenotyping is also examined in combination with active data gathering to detect the early onset of depression in healthy populations. A prospective study on the app "Warn-D", for instance, follows a cohort of students to 'warn' those who have a risk of being depressed, based on the app's analysis (Fried et al., 2023).

Given these current developments in mental health services and e-mental health, it is likely that digital phenotyping – and apps that operate solely on that basis – will be established as an alternative to standard screening tools for depression, both for clinical purposes in routine healthcare encounters and for direct-to-consumer use. While promising in terms of user friendliness and resource efficiency, the development of mental health apps that operate exclusively on digital phenotyping and gather data passively is connected with multiple epistemological and ethical questions that are part of an ongoing debate (Baumgartner, 2021; Coghlan & D'Alfonso, 2021; Stanghellini & Leoni, 2020; Tekin, 2020). Ethical considerations concerning e-mental health have generally focused on a variety of topics, such as autonomy (Schmietow & Marckmann, 2019), responsibility (Martinez-Martin & Kreitmair, 2018), and inequality (Skorburg & Yam, 2021). Another ethical topic that has recently received more attention in the use of machine learning is epistemic injustice (Hull, 2023; Pozzi, 2023a, b; Slack & Barclay, 2023; Symons & Alvarado, 2022).

We focus in this paper on apps that operate on passive data collection and digital phenotyping to screen for depressive episodes, in the following referred to as 'DP-apps.' We examine to what extent DP-apps generate cases of epistemic injustice by paying particular attention to their impact on epistemic agency in different contexts in which the apps can be used, including healthcare encounters and direct-to-consumer use. The paper draws on the existing literature on epistemic injustice in mental healthcare, scholarly and clinical discussions on the concept of mental illness, and the features of digital phenotyping and its underlying technological approach. Firstly, we will outline the conceptual background and further procedure, which includes three hypothetical scenarios that we use to illustrate and clarify our ethical analysis (2.). Subsequently, we argue that DP-apps have the potential to undermine different preconditions for epistemic agency and can lead to three different forms

<sup>1</sup> This model is characterized by its ability to combine information of different modalities (e.g., information contained in text and audio sequences on the user's state and complementary information that can also be part of the sequences). Moreover, it is considered to be a 'neural network' as it allows for nonlinear transformations of the input. Adapted from the human brain, the network model simulates the interconnections of data points (in the brain: the neurons) and increases its predictive abilities through learning and training. Artificial neural networks are a type of machine learning, which is a branch of AI (Sarker, 2021).

<sup>2</sup> A depressive episode is defined in the Diagnostic and Statistical Manual (DSM-5) (American Psychiatric Association, 2013, pp. 160–161) as a two-week period with at least five of the following symptoms and one of them being either depressed mood or loss of interest or pleasure: (1) depressed mood; (2) loss of interest or pleasure; (3) change in appetite or weight; (4) sleep troubles; (5) psychomotor agitation or retardation; (6) fatigue or loss of interest; (7) feelings of hopelessness or guilt; (8) concentration troubles; or (9) recurrent thoughts of death or suicide.

of epistemic injustice: hermeneutical (3.), testimonial (4.1) and contributory injustice (4.2). Throughout our analysis, we consider the influence of DP-apps on epistemic practices on an individual level, in healthcare settings, and on the structural level. Our overall objective is to make risks of epistemic injustice visible, promote sensitivity among all stakeholders, and develop recommendations for further development and implementation of this technology.

## Conceptual background and methodology

Theories of epistemic injustice focus on harms that affect people as ‘epistemic agents.’ The latter term refers to individuals who use shared epistemic resources to collect, generate, and distribute knowledge, justified belief, or understanding, and to revise existing epistemic resources if necessary (Dotson, 2014, p. 115). ‘Epistemic resources’ refer to language, concepts, theories, and standards of judgment that epistemic agents use to formulate propositions and make sense of social experiences (Pohlhaus, 2012). We consider DP-apps as ‘epistemic tools’ that users utilize for a specific epistemic task, i.e., to learn about their mental well-being by receiving information on their personal current risk of depression.

To be a successful epistemic agent, different preconditions need to be met, which can broadly be divided into aspects regarding epistemic resources and uptake. Firstly, a person relies on having fitting epistemic resources to make sense of their experiences and share them. Fricker, for instance, argues that the development of the concept of *sexual harassment* helped many women to understand and name a specific form of gendered violence at the workplace (2007).<sup>3</sup> Gaps in collectively shared epistemic resources can lead to hermeneutical injustice, i.e., a disadvantage in making sense of or communicating important social experiences due to structural marginalization in shaping epistemic resources. Secondly, a person’s testimony needs to receive adequate social uptake on both the interindividual level, which is especially important for healthcare encounters, and the structural level. In the case of testimonial injustice, a person’s testimony receives deflated levels of credibility based on the prejudice against their social identity.<sup>4</sup> Thirdly, contributory injustice, a concept introduced by Dotson (2012), arises if a speaker’s testimony is unintelligible because they use epistemic resources developed in marginalized

communities. If these resources are actively ignored by a listener, the latter is unable to understand the speaker’s contribution. Contributory injustice can affect people from marginalized communities whose epistemic resources are ignored by people outside the community due to societal power relations. It is often rooted in the lack of necessary social, political, and economic power to successfully spread their epistemic resources beyond their communities.

In the next chapters, we will analyze how DP-apps can impact each of these three preconditions, leading to hermeneutical (3.), testimonial (4.1), and contributory injustice (4.2). We argue, more specifically, that

**Hypothesis 1:** the use of DP-apps based solely on passive tracking generates or exacerbates hermeneutical injustice for users;

**Hypothesis 2:** the use of DP-apps may trigger testimonial injustice against users in healthcare encounters; and

**Hypothesis 3:** the use of DP-apps may increase the risk of contributory injustice for people with experience of depression.

Considering the broad range of mental illness, differences between severe and common mental health conditions, and the current state of technological development, we focus on one specific disorder, namely, a depressive episode. We use the term ‘user’ to designate people who experience mental health symptoms and use mental health services, such as local consultations and digital tools.

Methodologically, we introduce three different hypothetical case scenarios, telling the stories of Blake, Kay, and Noa, to clarify and exemplify specific points of our analysis.<sup>5</sup> While we think that the case scenarios reflect a range of plausible and relevant constellations from a perspective of epistemic injustice, they are not meant to encompass all possible and ethically interesting cases. The starting point for all scenarios is as follows: Blake, Kay, and Noa, who have no prior history of mental health concerns, download a free DP-app (that relies exclusively on passive data analysis) to track their mental health and risk of depression. They regularly use their smartphones for communication, information, and entertainment, while the DP-app runs in the background. Some weeks after they have downloaded the app, the following happens:

### *Scenario 1 – Blake is sad.*

Blake experiences a constant sadness, a loss of motivation and appetite, and problems with sleep that last more than two weeks. Blake stops leaving their bed and responds less to the messages of their friends. Blake begins to wonder whether they might be seriously depressed. Some days after

<sup>3</sup> Fricker takes the example from Susan Brownmiller’s memoir of the US women’s liberation movement (Brownmiller, 1990, as cited by Fricker, 2007, p. 148 ff).

<sup>4</sup> Testimonial and hermeneutical injustice are part of Fricker’s framework that she elaborated in “Epistemic injustice. Power and the ethics of knowing” (2007).

<sup>5</sup> The case examples are informed by clinical trial protocols (Leaning et al., 2024), the medical expertise of one of the authors, and similar case examples in conceptual discussions of depression, see Horwitz and Wakefield (2012, pp. 10–11).

having that thought, the app, by itself, issues a warning that Blake's symptoms could be the beginning of a depressive episode.

### ***Scenario 2 – Kay has a conflict.***

Kay is currently involved in a complex conflict with their partner, while preparing for an important exam. Kay feels sad, anxious, and stressed about the situation. To cope with the situation, Kay decides to focus on their work and to reduce contact with all the people involved in the conflict until they have time for it. They have less contact with friends and family, use their phone less, and sleep less, because they study in the evenings. Kay is satisfied with their coping strategy and feels confident that things will be OK after the exam. The app issues a warning that Kay could have a depressive episode.

### ***Scenario 3 – Noa feels alienated.***

Noa has felt sad, hopeless, numb, and alienated from their body for several weeks. Noa is a good student and manages to keep up their social life and study. They think they might have a depressive episode, something which they experience as a very profound and spiritual experience, even though painful, because they take this to be an occasion to think about their priorities and choices in life. Feeling hopeless, Noa thinks that no person could possibly guide them through this experience. The app does not issue a warning.

## **Resources-related aspects – DP-apps and hermeneutical injustice**

A first important aspect for a successful epistemic agency concerns the availability of fitting epistemic resources, such as concepts, theories, or images, to make sense of a social experience. As Fricker points out, epistemic agents can be harmed if adequate epistemic resources are not collectively available due to social power relations within a society: “hermeneutical injustice occurs [...] when a gap in collective interpretive resources puts someone at an unfair disadvantage when it comes to making sense of their social experiences” (2007, p. 1). If structural power relations negatively affect the availability or adequacy of epistemic resources that people need to make sense of or communicate their experiences, for example, concepts of illness, their experiences are left ‘obscured’ or ‘badly understood.’ This disadvantage is unfair if it originates from what Fricker calls “hermeneutical marginalization,” i.e., when an individual or a group is excluded from participating equally in significant epistemic practices (e.g., education, science, politics, media) and contributing to collectively shared epistemic resources (2007, p. 152). Fricker gives the example of a woman who suffers from a depressive episode after giving birth to her son (2007, p. 148). Lacking the concept of postpartum

depression, which is not part of collectively shared concepts in her society, she fails to make sense of her experience. Instead, she feels guilty about her alleged *personal* deficiencies. Since postpartum depressions are, with a prevalence of around 17%, very common among mothers and have a complex bio-psychosocial etiology (Shorey et al., 2018), the woman in Fricker's example, is disadvantaged in understanding what she is going through. According to Fricker, her misinterpretation can be traced back to the fact that women and people with mental illness cannot participate equally in the epistemic practices that build and shape concepts of women's mental health within a patriarchal culture, and are, thereby, hermeneutically marginalized. Cases of hermeneutical injustice are, thus, characterized by three central features: (1) gaps in collectively shared epistemic resources (2) result in significant disadvantages in making sense of or communicating a social experience, (3) whereby the gaps are caused by the hermeneutical marginalization of those who are having this social experience, based on power structures. In the following, we argue that the use of DP-apps, because of their functionality, generates or promotes cases of hermeneutical injustice. We start by (1) describing hermeneutical gaps engendered by a bias towards biostatistical understandings of mental illness built into the functionality of DP-apps, and proceed by (2) discussing these gaps in relationship to the disadvantages they cause in users when it comes to making sense of their experiences. (3) We suggest that these disadvantages derive from the hermeneutical marginalization of people with depression in the apps' development, and further contribute to their hermeneutical marginalization.

(1) Hermeneutical gaps caused by DP-apps stem from their technological approach, which builds a bias towards a biostatistical understanding of depression into the app's functionality. In the literature, different forms of bias in training datasets and algorithmic functionality that distort predictions systematically have been described (Alvarado & Morar, 2021; Klugman, 2021; Müller, 2021). In addition to well-documented artificial intelligence (AI) algorithm biases that lead to erroneous or discriminatory analyses when it comes to marginalized groups (Norori et al., 2021), we want to highlight a more profound, conceptual bias, which would also arise in the case that representative and unbiased datasets were available. The technological approach is not neutral in theoretical terms, since the very idea of detecting depressive episodes via passive data gathering and digital phenotyping presupposes a biostatistical account of disease. Proponents of such accounts argue that disease can be defined objectively by referring to states of normal functioning based on statistical normality within the relevant reference class (Boorse, 1977). Similarly, on that account, mental disorders describe states of (e.g.,

behavioral, emotional, and cognitive) functioning below a ‘normal’ threshold compared to the reference class, independently of social norms and values. In this vein, passive data gathering uses objectively measurable and statistically processable data of a person’s observable functioning and compares it to the profile of other users to predict the onset of a depressive episode. Thus, it is built into the app’s underlying functionality that detecting a depressive episode is based on measurable statistical deviance.

Some depressive episodes involve symptoms that correspond to deviances in functioning that are covered by the app’s measurements, such as Blake’s behavioral changes (staying in bed, having fewer social contacts). In fact, a study by Moshe et al. (2021), that investigated the use of passive data in digital phenotyping to predict changes in depressive and anxiety symptoms, found statistically significant, but weak correlations between time in bed and depression or the variability of locations visited and depression. However, depressive episodes frequently involve feelings of pain, fatigue, burning, tension, numbness, or heaviness (Fuchs, 2013). These symptoms are better described through phenomenological approaches (e.g., Carel, 2016; Fuchs, 2013; Wardrope & Reuber, 2022). Phenomenological approaches to mental illness support that illness cannot be understood using biostatistical models because illness inherently involves changes in how the person relates to their body, and their social and physical world (Carel, 2007). In keeping with this, Fuchs (2013) argues that depression is more accurately understood as a change in a person’s ‘interactivity’ and their ‘intercorporeality,’ i.e., the way a person relates to their lived body and their interpersonal environment, rather than changes in individual functioning. This is particularly apparent in Noa’s case. Significant dimensions of Noa’s suffering, such as the feeling of alienation and numbness, are not detectable in their digital phenotype, because they do not translate into measurable changes. In addition, social constructivist approaches to depression assume that the definition of mental disorders inherently depends on social practices and value judgments; proponents stress that mental disorders cannot be defined or identified without referring to social factors (Horwitz, 2012; van Riel, 2016). Horwitz and Wakefield (2012) argue that not considering the social context of a change in functioning may lead to the pathologizing of behavioral or emotional responses to stressful social situations. It seems that DP-apps, by neglecting the social context, risk misidentifying states as pathological that are norm-deviant, but otherwise unproblematic. Kay’s distress, for instance, is caused by a social conflict and the important exam they have to prepare for. Their decision to concentrate on their exam and postpone solving the conflict appears as a useful coping strategy to an external stressor. It is conceivable that after the exam and the resolution of the conflict,

Kay will be emotionally recovered. Some would, therefore, argue that diagnosing a depressive episode in Kay would mean to medicalize a social or interpersonal problem, and that this would be unwarranted (Horwitz & Wakefield, 2012).

Consequently, biostatistical accounts of mental disorders have been criticized to fail to acknowledge the importance of subjective meaning-making, value judgments, social norms, and structural contexts in deciding which condition qualifies as a mental disorder (Carel, 2007; Kingma, 2007). DP-apps do not require any active user input or value judgments. First-person evaluations and social context are not clearly related to a person’s observable behavior. Uusitalo and colleagues caution against the use of AI for detection and diagnosis in psychiatry because “[l]eaving aside salient features of subjective experience and social factors runs the risk of simplifying the categories to the extent that the phenomenon is misconstrued” (2020, p. 3). Consequently, DP-apps, by design, neglect social, structural, and phenomenological aspects of depression. DP-apps can detect cases of depression corresponding to the biostatistical model of depression, and systematically miss other forms of depression. Additionally, DP-apps may wrongly detect non-pathological states as depressive, because they fail to acknowledge the social context and a person’s own interpretation of a situation. DP-apps, understood as epistemic tools which users utilize to make sense of their social experiences, engender hermeneutical gaps and may produce misinterpretations.

(2) This is epistemically disadvantageous for users, because misinterpretations caused by DP-apps may complicate a user’s understanding of their lived experiences, depending on how much users rely on the app. In addition to the hermeneutical gaps that result from the design of DP-apps, AI-based systems are characterized as widely opaque black boxes, which is a major issue in AI or (big) data ethics (Müller, 2021). Opacity exacerbates the disadvantages in understanding, i.e., contextualizing and evaluating the app’s output. Opacity implies that a system’s output, even if it accurately predicts a certain outcome, remains unexplained and (perhaps) unexplainable, to the extent that the result is generated by machine learning algorithms (Alvarado & Morar, 2021; Burrell, 2016; Klugman, 2021; Theunissen & Browning, 2022; Zarsky, 2016).

Other authors have argued that the opacity of AI systems may reinforce epistemic injustices that are caused by such systems (El Kassar, 2022; Symons & Alvarado, 2022). El Kassar’s (2022) analysis of AI systems for automatic gender recognition highlights that the system’s misinterpretations of a person’s gender can lead to self-doubt in an affected individual. The black box problem exacerbates this self-doubt by making it difficult for individuals to understand on which features their gender was misinterpreted. The

problem of opacity is relevant to both the users affected and IT experts who are incapable of sufficiently understanding how a particular machine learning algorithm based on a specific dataset generates a certain output.

Similarly, it remains unclear in machine learning based DP-apps how exactly the app generated a certain output; neither Blake, Noa, nor Kay might be able to sufficiently understand what patterns the algorithm used to predict the risk of a depressive episode. Medical and technical aspects that imply presuppositions (or defaults) about ‘normal’ digital interaction (such as the underlying model of mental illness, its statistical conceptualization, and digital implementation) are difficult to assess from a lay perspective. The app’s opacity makes it harder to gain a critically reflected understanding of one’s mental health issues in comparison to someone who has access to qualified human mental healthcare professionals who can be asked to make their criteria and procedures transparent and meet the individual needs for information and support. DP-apps, as an epistemic tool, can compare an individual’s digital phenotype to a statistical pattern. However, this does not (yet) provide a meaningful explanation and justification for a generated output with potentially far-reaching consequences for the user affected. This may be connected to further behavioral or cognitive changes that may also be disadvantageous for users. In the case of a positive result, Kay, who believes they are mentally healthy, is likely to be confused, unsettled, or skeptical about receiving a warning about depression. To gain a deeper understanding, Kay needs to know more than that there are (clinically relevant) signs of depression in their digital phenotype. Starting to reflect upon the result, they also need to know how and why certain aspects of their behavior are deemed pathological.

We have argued so far that misinterpretations in the usage of DP-apps are possible because DP-apps as epistemic tools involve hermeneutical gaps. Bias in the underlying technological and theoretical approach puts app users at a disadvantage in understanding their mental issues, which is exacerbated by opacity. This is especially relevant for people who lack psychiatric expertise or personal experience with mental illness, have difficulty accessing alternative mental healthcare services, suffer from severe symptoms that significantly impair self-perception, or are particularly vulnerable to epistemic harm or social stigma because of marginalized group membership. These preconditions make it more difficult for people to contextualize the app-generated outputs, evaluate them against the background of their own experience and knowledge, and, if necessary, scrutinize them critically.

(3) We will now argue that these hermeneutical gaps stem from the hermeneutical marginalization of the people affected and that the gaps can also reproduce this

hermeneutical marginalization. This necessitates considering the conditions under which these apps are developed, taking into account social power dynamics. The development of DP-apps rests on the idea that the detection of depression is possible with an algorithmic system that relies on a biostatistical understanding of mental disorders. Feeding the model with passively gathered data via digital phenotyping is considered to be an adequate instrument for recognizing complex mental phenomena, such as depression. This reliance on data collection as the best means of approaching the world has been dubbed ‘data fundamentalism,’ according to which “massive data sets and predictive analytics always reflect objective truth” (Baumgartner, 2021, p. 6). Data fundamentalism is closely related to biomedical approaches within contemporary psychiatry, which, despite the advocacy for more social, relational, or phenomenological approaches, remain dominant in psychiatric research.<sup>6</sup> It has been argued that psychiatric research based on a biostatistical understanding excludes service users and perceives them as mere objects of research. They are not actively, as epistemic agents, participating in shaping diagnostic categories, understandings of mental illness, and the development of study interests and designs (Bueter, 2019; Crichton et al., 2017; Kidd & Carel, 2018; Kidd et al., 2022; Lakeman, 2010; Miller Tate, 2019). This trend is reproduced in the development and implementation of DP-apps: none of the studies included in the systematic review on digital phenotyping to predict the onset of a depressive episode used participatory methods (Leaning et al., 2024), which are perceived as an adequate and efficient means of allowing users to contribute to the research as epistemic agents, fostering both social and personal transformation (Thomas et al., 2023). Accordingly, the underlying model and the design of studies on DP-apps are indicative of users’ hermeneutical marginalization in the development of DP-apps.

Additionally, we assume that the implementation of DP-apps can further reinforce the hermeneutical marginalization of service users. Given that DP-apps are based on a biostatistical understanding of depression that does not rely on users’ subjective explanations, value judgments, and experiences, their implementation may further strengthen the trend of marginalizing users’ experiences of depression: if their accounts of depression are irrelevant to identifying depression, their value for psychiatric research may be misjudged and their voices may be overlooked in healthcare encounters. We conclude that there are cases in which people suffer hermeneutical injustice by using DP-apps as epistemic tools because (1) of the gaps in the epistemic resource that (2) put users at a disadvantage when it comes

<sup>6</sup> This is, for instance, reflected in the funding policies of the US National Institute of Mental Health, which has spent a large majority of funding on neuroscientific basic research (Torrey et al., 2021).

to making sense of their experiences (3) in the context of users' hermeneutical marginalization within contemporary psychiatric research.

## Uptake-related aspects

### Epistemic agency and DP-apps in healthcare encounters

In addition to having fitting epistemic resources, the social uptake of the testimony in question is necessary for sharing knowledge successfully. As Dotson (2011, p. 237) points out, a successful linguistic exchange does not only depend on the speaker's epistemic and communicative skills but also on the hearer's acknowledgment of a person's speech act as it is meant to be taken. When a person testifies knowledge, it can only be successful if they are recognized as a knower and their communication is recognized as a testimony. The testimony must, therefore, receive appropriate social uptake as being credible. Uptake can be impacted on two levels: firstly, on the individual level, if a speaker is not acknowledged as a knower, and secondly, on the structural level, if the epistemic resources of a social group are not acknowledged as such.

Starting with the first, Fricker (2007) notes that people constantly, and often unconsciously, check a speaker's credibility, i.e., their honesty and competency, in linguistic exchange. Linguistic exchange is located in a context of social power relations which shape credibility assessments. Fricker speaks of "testimonial injustice" (2007, pp. 28 ff) if prejudice based on a speaker's social identity unduly influences the listener's perception, leading to a credibility deficit. In the following, we aim to argue that DP-apps can trigger testimonial injustice in healthcare settings, especially in GPs' offices.

We first consider the baseline situation in which users and healthcare professionals interact. It has been argued that people who show or report psychiatric symptoms are particularly vulnerable to experiencing testimonial injustice from healthcare professionals during healthcare encounters based on negative stereotypes about mental illness (Crichton et al., 2017; Drożdżowicz, 2021; Faissner et al., 2022; Kurs & Grinshpoon, 2018; Scrutton, 2017). Fricker (2007) suggests that in our everyday communicative practices, credibility judgments are often made spontaneously in our automatic processing of information. Pozzi (2023b, p. 538) further argues that "credibility assessments are particularly prone to be distorted by biases and stereotypes connected to a person's social identity because to form these, we usually rely on so-called *markers of trustworthiness*." That means, people rely on markers that identify a person as a reliable

knower to assess whether believing a person will generate knowledge, and stereotypes are often used as markers of trustworthiness.

Empirical evidence indicates that people are biased against those with depressive symptoms due to stereotypes on mental illness. Negative stereotypes, such as 'mad,' 'crazy,' 'incoherent,' 'delusional,' or 'cognitively disturbed,' are common among the general population and remain stable over time (Angermeyer et al., 2013; Thornicroft et al., 2007; von Kardoff, 2017).<sup>7</sup> The same appears to hold true for healthcare providers: According to a study investigating implicit biases and depression, internal medicine residents showed a significant association between depression, negative attitudes, and ideas about the uncontrollability of depression (Crapanzano et al., 2018). The literature on stereotyping and implicit biases indicates that these have important effects on clinical practice, including the choice of further diagnostic steps and treatment decisions (Puddifoot, 2019).

Being stereotyped and disregarded in healthcare settings is a common experience for people with mental illness. One example is that users with mental health diagnoses often face difficulties in making their somatic health complaints or drug-related side effects heard by their physicians (Faissner et al., 2022; Golomb et al., 2007). Mental health diagnoses strongly shape how a person is perceived in healthcare encounters: A qualitative study revealed that in the presence of a mental health diagnosis, medical staff tend to attribute all symptoms to that diagnosis, so that other relevant diagnoses are missed, including bone fractures (Shefer et al., 2014).<sup>8</sup> Additionally, it is a special feature of psychiatric interviews to interpret "patient's statements as manifestations of illness" (Sakakibara, 2023, p. 490). More specifically, not being able to feel one's own sadness is assumed to be a possible symptom of depression. According to the DSM-5, an influential diagnostic manual in mental healthcare, "in some cases, sadness may be denied at first but may subsequently be elicited by interview" and, furthermore, "in some individuals who complain of feeling 'blah,' having no feelings, or feeling anxious, the presence of a depressed mood can be inferred from the person's facial expression and demeanor" (APA, 2013, p. 163). Thus, the DSM-5 seems to encourage

<sup>7</sup> Stereotypes about mental illness often intersect with negative stereotypes shaped by racism, ableism, or cis-heterosexism. Therefore, it is important to consider the intersectional disadvantage that service users could encounter who are marginalized regarding multiple aspects of their social identity.

<sup>8</sup> A psychiatric nurse participating in the study related the story of a "behaviorally disturbed" person who was screaming in the emergency room. The person was assumed to be psychotic and was "medically cleared" by the emergency staff. Noticing a swollen foot, the nurse organized an X-ray which showed a fractured foot (Shefer et al., 2014).

medical staff to question patients' testimonies, including denials of sadness, and to privilege externally observable signs of depression over a person's self-report. DP-apps, thus, come into play in a baseline "credibility economy" (Fricker, 2007) in which service users may be perceived as unreliable givers of knowledge, and their testimony is under special scrutiny.<sup>9</sup>

In the following, we will examine how DP-apps may influence this epistemic baseline situation. Pozzi (2023b) argues that in medical contexts, AI-based algorithmic risk scores may be treated as markers of trustworthiness to assess the credibility and relevance of a user's testimony, especially in cases in which the symptoms reported by a patient are not related to an easily detectable source (i.e., symptoms that cannot easily be traced back to a specific medical test vs. patient reports pain and X-ray shows a broken bone). Baumgartner suggests: "self-tracking from digital devices is framed as providing trustworthy data in contrast to the individual body's perception which is marked as untrustworthy or at least not reliable enough to be the sole ground on which diagnosis should be based" (2021, p. 6). The output of DP-app's may be treated as a marker of trustworthiness of a users' credibility because it could be interpreted as objective evidence, comparable to a diagnostic test with a certain reliability. This idea is supported by the literature suggesting that medical professionals exercising under the biomedical paradigm are trained to privilege data over service users' testimonies (Drożdżowicz, 2021; Kidd & Carel, 2018; Slack & Barclay, 2023).

In order to substantiate this claim, it is helpful to further consider the context in which DP-apps are implemented. A meta-analysis on the diagnosis of depression in primary care suggests that GPs may not be adequately trained to detect and diagnose depression and often fail to do so (Mitchell et al., 2009, 2011). In fact, GPs were *more* likely to *falsely* diagnose a person with a moderate or severe depressive episode than to correctly diagnose depression: "for every 100 unselected cases seen in primary care, there are more false positives (n=15) than either missed (n=10) or identified cases (n=10)" (Mitchell et al., 2009, p. 609). General practitioners who feel insecure in diagnosing depression may welcome external evidence perceived as objective, and may treat the app's output as a marker of trustworthiness. However, as argued above, DP-apps are not neutral as they privilege a narrow understanding of depression, based on

<sup>9</sup> This is not to say that mental health professionals *intentionally* harm service users epistemically or *actively* aim to downgrade their testimony. As Saul (2017) suggests, prejudice based on stereotypes often interacts with a person's credibility judgment through implicit biases which are even often opposed to a person's conscious beliefs. Therefore, a physician could hold the conscious belief that users with depression are equally reliable knowers, while still downgrading their credibility.

dysfunction and statistical abnormality. Thus, DP-apps can be considered unreliable markers of trustworthiness. Starting from this baseline situation, we will now assess two different constellations to analyze how DP-apps may interact with users' credibility assessments and epistemic agency: (1) the interpretation from the app and the user misaligns, and (2) the interpretation of the app and a user aligns.

(1) In the case of Kay, there is a misalignment between the app's and Kay's interpretation of the situation: Kay thinks that they are okay, while the app issues a warning that Kay may have a depression. Such cases can be understood as 'false positives,' i.e., an identification as pathological of a non-pathological state. If Kay decides to visit a GP, we see a considerable risk that the GP may unduly privilege the app's output over Kay's testimony, and that Kay may suffer testimonial injustice. The reason for this claim is that implicit biases on depression, insecurities in diagnosing depression, and a general bias towards objectifiable tests in medicine all raise the epistemic weight of the app's output at the expense of Kay's credibility. Additionally, based on the assumption that people with depression might deny sadness, a GP may interpret Kay's assertion to be okay as a symptom of depression. Yet, in such a case, the GP would fail to acknowledge that users, such as Kay, have direct access to their own experiences and are thereby epistemically privileged (Drożdżowicz, 2021). The GP would, thus, treat Kay as a passive 'source of information,' rather than an active informant, i.e., a *subject* of knowledge, and, thereby, wrong them (Sakakibara, 2023).<sup>10</sup>

This claim can further be empirically substantiated by findings from qualitative research: The authors in a study on GP treatment recommendations in mental health conditions found that even though three-quarters of the patients initially resisted treatment recommendations, most patients were prescribed medication or referred to talking therapy (Ford et al., 2019). This might indicate a possible trend of disregarding users' concerns in primary care. In a context where GPs are negatively biased against users with mental

<sup>10</sup> At this point, someone could object that it is, in some cases, warranted to downgrade the credibility of a person showing depressive symptoms, as a depression might be accompanied by cognitive biases. The latter, such as ruminations or overgeneralizations, could, in fact, impair a person's epistemic competency, so that a downgrading of their credibility could be warranted. While having a depressive episode constitutes a risk factor for reduced epistemic competency, the presence of a depressive episode is never, on its own, a sufficient reason to downgrade a person's credibility. Instead, the competency of a speaker is to be assumed, and only to be probed in a healthcare encounter if in doubt. In this case, examining a person's epistemic competency is ethically warranted, for instance, to ensure an informed consent standard (for a discussion, see Talukdar, 2021). However, in the case of testimonial injustice, the downgrading of a person's credibility is based on prejudice, not on actual evidence of reduced epistemic competency or sincerity.



health symptoms based on structural stigma and are not well trained to diagnose depression, DP-apps are likely to be treated as markers of trustworthiness, and may thus influence the ‘credibility economy’ to the disadvantage of users. Testimonial injustice is, thus, likely to occur in the context of false positives when a user denies being depressed.

The situation is different in the case of Noa: The app does not issue a warning, even though Noa is suffering and might benefit from mental health support. Noa’s case can be described as a ‘false negative.’ Based on our prior argument and the fact that GPs are likely to wrongly identify depression in non-depressed people, we assume that if Noa visited a GP and told them that they felt depressed, the GP would take their complaints seriously. Yet, the app may, nevertheless, influence Noa’s epistemic agency. If Noa expects the app to confirm their assumption that they are depressed and the app does not issue a warning, Noa may question their own intuition and be discouraged from seeking social or medical support, as argued above. This requires further research focusing on the interaction between apps and users’ epistemic agency.

(2) In cases of alignment between the app and a user, it seems *prima facie* unproblematic if a GP takes the output from the app into account. The app provides information that a user cannot provide by themselves, i.e., the analysis and statistical comparison of a users’ digital phenotype. The app’s output seems comparable to external evidence (such as medical test results) that may help a GP to build an informed clinical judgment. Arguably, DP-apps may, nevertheless, support epistemically unjust behavior. We usually accept a person’s account of their physical and mental well-being, respecting their epistemic authority and granting them an epistemic privilege regarding bodily sensations and experiences. However, based on the app’s functionality, this information is not relevant to detecting depression (see the discussion of hermeneutical injustice), therefore, the app’s functionality on a meta-level justifies discarding these forms of testimony. Thus, focusing on the app’s analysis, even if it offers additional information, instead of focusing on a person’s testimony in healthcare encounters may strengthen both the trend towards data fundamentalism described above and negative prejudices about the epistemic competency of people who report mental health complaints (Slack & Barclay, 2023).

To sum up, DP-apps may promote unintentional epistemically unjust behaviors by third parties, especially by GPs, which enhances the risk of testimonial injustice against users. This may have negative consequences on users’ epistemic agency and confidence in their judgments and increase the risk of inappropriate healthcare. As Fricker (2007, pp. 48 ff) explains, systematically experiencing testimonial injustice can lead to a person losing confidence

in their belief or its justification, so that the belief no longer fulfills the conditions for knowledge, or they can lose confidence in their intellectual abilities. That person might come to view themselves as an unreliable knower. Kay, for instance, could experience reduced epistemic self-confidence and agency if their testimony systematically receives less attention and interest due to the focus on the information provided by their DP-app in the healthcare encounter. These effects can also occur outside the healthcare setting in the everyday context of user-app interaction. Receiving a warning, while not being depressed, could influence Kay’s interpretation and behavior. Wyatt (2018) speaks in the context of health apps and digital healthcare about ‘apptimism,’ “an uncritical, implicit trust in apps” based on the fact that “most of us carry and use our smartphone all day, so we trust everything it brings us.” While the use of DP-apps may influence a person’s confidence in their judgments on an individual level, DP-apps may also influence epistemic practices on a structural level. We will consider this in the next section.

### DP-apps and social uptake – contributory injustice

Dotson (2012, pp. 33 ff) introduces another form of epistemic injustice in her critical discussion of Fricker’s work which touches both the individual and structural level of epistemic injustice, so-called ‘contributory injustice.’ The focus here lies on the social uptake of epistemic resources developed by marginalized groups on the structural level. While many analyses of mental healthcare practices focus on testimonial and hermeneutical injustice (e.g., Crichton et al., 2017; Kurs & Grinshpoon, 2018), contributory injustice has been less discussed so far (e.g., Miller Tate, 2019). However, the concept is similarly important for understanding and carefully analyzing epistemic injustices in mental healthcare.

Dotson (2012) points out that in the case of gaps in collectively shared epistemic resources, people in marginalized social positions often develop epistemic resources to make sense of their social experiences and reality which knowers from dominant social positions do not share. Consequently, different sets of epistemic resources are socially available to make sense of a given social experience, where the resources developed by affected people would be more appropriate to capture it. Thus, Dotson’s definition of contributory injustice allows one to recognize the epistemic resources and agency of marginalized users. Contributory injustice arises if knowers in dominant social positions ignore, distort, or disregard these epistemic resources due to willful ignorance. In such cases, the use of distorted or inadequate epistemic resources by the dominantly positioned listener prevents the successful social uptake of a person’s testimony, even if the

speaker uses fitting concepts, so that their epistemic agency is undermined. As Dotson summarizes, contributory injustice arises in such situations where “an epistemic agent’s willful hermeneutical ignorance in maintaining and utilizing structurally prejudiced hermeneutical resources thwarts a knower’s ability to contribute to shared epistemic resources within a given epistemic community by compromising her epistemic agency” (2012, p. 32).

The following example will serve as an illustration of Dotson’s concept. While the biomedical understanding of mental health crises in terms of the DSM-5 is the standard approach which structures most medical interactions and social practices (e.g., covering treatment costs and sick leave), this is far from being the only approach to make sense of such experiences. Several groups of mental health activists and service users have developed various concepts that differ importantly from the medical model (Hoffman, 2019; Radden, 2012; Rashed, 2019). Some of these activist groups refer to themselves as Mad Pride, in reference to other social justice movements such as Gay Pride, and suggest that framing their lives with mental diversity as deficient constitutes an unwarranted epistemic impoverishment. Instead, as the Icarus Project, one of the Mad Pride self-help activism networks, suggests, differences in mental experiences and functioning should be understood as “dangerous gifts” which need “cultivation and care,” as they can be a source of both creativity and inspiration as well as hardship and suffering (DuBrul, 2014, p. 266). In their view, the medical model reduces mental disorder to dysfunction and deficiency, which need to be eliminated. This subverts service users’ possibility of living meaningful lives with mental diversity, in which positive aspects of mental diversity can be acknowledged as a valuable part of their identity, while aspects of mental diversity which cause suffering receive adequate care. Hence, it should be considered that experiences of mental distress can be interpreted variously by different people, which leads to diverse coping strategies.

The discussion of the Mad Pride approach exemplifies that different epistemic resources, in terms of concepts, theories, and interpretations, are available to refer to mental health crises or mental diversity, beyond the biomedical approach. While studies recently started to involve users in the development process, for instance, by integrating them into app design phases (see study protocol by Young et al., 2022), the apps, so far, do not consider alternative approaches to mental health, to the best of the authors’ knowledge. This can have important effects not only on epistemic practices within healthcare encounters but also on a structural level. We can consider mental health apps as epistemic tools which, by their presence and use, influence and shape the epistemic resources that are collectively available in society. They have the potential to support specific epistemic

attitudes in the general public and mental health professionals, for instance, epistemic attitudes that favor biomedical interpretations of depressive symptoms, and neglect others, such as the Mad Pride movement or phenomenological approaches (for an overall discussion of this imbalance in mental health, see Huda, 2021; Kiesler, 1999).

In one of our case scenarios, Noa experiences their time of sadness as a spiritual and painful crisis which was important to their life and, therefore, nothing they would want to miss. However, their report risks not receiving an adequate uptake. If a listener relies on an account of mental illness according to which the latter is a pathology which is always a ‘bad thing’ to have, then Noa’s account of attributing meaning to their crisis remains unintelligible. DP-apps may thus contribute to shaping an epistemic landscape in which it becomes more difficult for people with depressive symptoms to share their experience if this experience does not align with the dominant mental disorder approach. Alternative interpretations and accounts of mental health crises may become inexpressible. The apps’ use may, thereby, contribute to undermining the epistemic agency of users with depressive symptoms, reinforce the marginalization of epistemic resources of non-dominant social groups and perpetuate contributory injustice.

## Conclusions

Epistemic injustice is currently an under-researched issue in digital mental health. We have applied theories of epistemic injustice, social epistemology, and the philosophy of psychiatry to analyze questions of epistemic injustice in using passive data gathering and digital phenotyping as standalone technology to predict the onset of a depressive episode. Focusing on three scenarios, the analysis has revealed how a user’s epistemic agency – a significant part of that person’s humanity – can be harmed by three different forms of epistemic injustice. We have argued that functional principles of digital technologies and theoretical presumptions about key concepts in psychiatry interdepend, and that social power relations can be hidden in and exacerbated by digital artefacts. If DP-apps become a standard in mental healthcare, comparable to 24 h electrocardiograms in cardiology, this could reinforce the trend toward approaches that stress objectifiable data and third personal knowledge in contrast to subjective firsthand reports. Such new standards of evidence could amplify the existing risk for mental health service users to suffer epistemic injustices.

We suggest that DP-apps have the potential to shape our epistemic landscape and shared epistemic resources and, thereby, lead to the further marginalization of models and approaches to mental health that do not correspond to the

biostatistical understanding of mental illness. More precisely, we have argued that

- 1) the use of DP-apps based solely on passive tracking generates or exacerbates hermeneutical injustice by involving a bias towards a biostatistical model of depression which is exacerbated by the problem of opacity in machine learning algorithms;
- 2) testimonial injustice can be triggered by the app in a social context in which GPs lack skills and knowledge on depression, and have negative implicit biases about people with depressive symptoms, especially in cases in which patients' testimony and the apps' output misalign; and
- 3) contributory injustice may be supported through the use of DP-apps because it supports biostatistical accounts of depression and marginalizes other approaches, thereby, raising the likelihood that a listener of a user's testimony ignores alternative approaches to depression and the epistemic resources used by the user to communicate their experiences.

DP-apps are still under development. Thus, our work provides normative considerations *before* a technology is widely implemented. Given the mental health diagnosis gap, the lack of skilled healthcare personnel, and the persisting stigmatization of mental illness, such apps have the potential to be improved and considered to alleviate this tight situation. We, therefore, encourage stakeholders on all levels to recognize the dimension of epistemic risks and incorporate measures into the apps' further research, development, and deployment to reduce these risks.

Russo et al. (2023) recently proposed a novel framework to examine the development of responsible AI. Their approach aims to bridge the divide between epistemic questions pertaining to AI, such as transparency and explainability, and ethical questions, such as how values and other normative considerations, for example, vulnerabilities, are implemented. They propose two practical strategies to improve AI development and assessment: focusing on the process rather than the output, and turning to more inclusive forms of assessment rather than pure expert assessment. Following up on their suggestion, it seems beneficial to involve users in the various phases of psychiatric and technological research, development, and design. Community-based participatory research seems a good avenue for future app development (Roberts, 2013). Such projects involve the prospective users of a technology actively as researchers and developers and center their interests and insights, striving for collaboration on an equal footing between academic researchers and users.

We suggest, based on our analysis, that apps which are solely based on passive self-tracking should not be used in the general population to predict the onset of depression, but if used, passive self-tracking should be combined with further tools for user-app interactions. It seems important that the app, by design, solicits the user's interpretation of their situation including their social context. Furthermore, DP-apps should be developed in line with different approaches to depression, including phenomenological, social, and activist approaches. Such approaches may be implemented through more holistic questions and interaction possibilities, the provision of supplementary information on different approaches to mental health and depression, and tools for community building.

From a structural perspective, exclusionary practices in mental healthcare, research, and development should be further examined and abolished, thus, preventing unjust epistemic power relations from being perpetuated or reinforced by digital artefacts. Our epistemic practices must progress and become more inclusive, appreciative, and emancipatory – going hand in hand with technological developments.

#### Abbreviations

|         |                                                               |
|---------|---------------------------------------------------------------|
| AI      | Artificial Intelligence                                       |
| DSM     | Diagnostic and Statistical Manual                             |
| DP-apps | Digital phenotyping apps for the self-detection of depression |
| GP      | General Practitioner                                          |

**Acknowledgements** The authors are very grateful toward the German Academy of Ethics in Medicine (AEM) for being awarded the Young Talent Award (Nachwuchspreis) 2022. The authors would like to thank the members of the working group Ethics in Psychiatry (Ethik in der Psychiatrie) of the AEM, the members of the Department of Practical Philosophy at the University of Greifswald, and the participants of the Colloquium "Neuroethics and Ethics in AI" hosted by the Institute of Neurosciences and Medicine, Forschungszentrum Juelich GmbH for giving critical and helpful feedback on earlier drafts of this article.

**Author contributions** All authors contributed to the study conceptualization. Eva Kuhn originally had the idea for the article; Mirjam Faissner and Sebastian Laacke performed a large part of the literature search and philosophical elaboration. Regina Müller particularly contributed to the original draft preparation and critically reviewed the line of arguments. All authors reviewed and edited the article.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

#### Declarations

**Competing interests** The authors declare they have no financial or nonfinancial conflict of interest. No funds, grants or other support was received.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Al Hanai, T., Ghassemi, M., & Glass, J. (2018). Detecting depression with audio/text sequence modeling of interviews. *Interspeech*, 1716–1720. <https://doi.org/10.21437/Interspeech.2018-2522>.
- Alvarado, R., & Morar, N. (2021). Error, reliability and health-related digital autonomy in AI diagnoses of social media analysis. *American Journal of Bioethics*, 21, 26–28. <https://doi.org/10.1080/15265161.2021.1926589>.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.
- Angermeyer, M. C., Matschinger, H., & Schomerus, G. (2013). Attitudes towards psychiatric treatment and people with mental illness: Changes over two decades. *British Journal of Psychiatry*, 203, 146–151. <https://doi.org/10.1192/bjp.bp.112.122978>.
- Babrak, L. M., Menetski, J., Rebhan, M., Nisato, G., Zinggeler, M., Brasier, N., Baerenfaller, K., Brenzikofer, T., Baltzer, L., Vogler, C., Gschwind, L., Schneider, C., Streiff, F., Groenen, P. M. A., & Miho, E. (2019). Traditional and digital biomarkers: Two worlds apart? *Digital Biomarkers*, 3, 92–102. <https://doi.org/10.1159/000502000>.
- Baumgartner, R. (2021). Precision medicine and digital phenotyping: Digital medicine's way from more data to better health. *Big Data & Society*, 8. <https://doi.org/10.1177/20539517211066452>.
- Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, 44, 542–573.
- Bueter, A. (2019). Epistemic injustice and psychiatric classification. *Philosophy of Science*, 86, 1064–1074. <https://doi.org/10.1086/705443>.
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3. <https://doi.org/10.1177/2053951715622512>.
- Carel, H. (2007). Can I be ill and happy? *Philosophia*, 35, 95–110. <https://doi.org/10.1007/s11406-007-9085-5>.
- Carel, H. (2016). *Phenomenology of illness*. Oxford University Press.
- Coghlan, S., & D'Alfonso, S. (2021). Digital phenotyping: An epistemic and methodological analysis. *Philosophy & Technology*, 34, 1905–1928. <https://doi.org/10.1007/s13347-021-00492-1>.
- Crapanzano, K., Fisher, D., Hammarlund, R., Hsieh, E. P., & May, W. (2018). An exploration of residents' implicit biases towards depression – a pilot study. *Journal of General Internal Medicine*, 33, 2065–2069. <https://doi.org/10.1007/s11606-018-4593-5>.
- Crichton, P., Carel, H., & Kidd, I. J. (2017). Epistemic injustice in psychiatry. *BJPsych Bulletin*, 41, 65–70. <https://doi.org/10.1192/pb.bp.115.050682>.
- Dotson, K. (2011). Tracking epistemic violence, tracking practices of silencing. *Hypatia*, 26, 236–257.
- Dotson, K. (2012). A cautionary tale: On limiting epistemic oppression. *Frontiers: A Journal of Women Studies*, 33, 24–47. <https://doi.org/10.5250/fronjwomestud.33.1.0024>.
- Dotson, K. (2014). Conceptualizing epistemic oppression. *Social Epistemology*, 28, 115–138. <https://doi.org/10.1080/02691728.2013.782585>.
- Drożdżowicz, A. (2021). Epistemic injustice in psychiatric practice: Epistemic duties and the phenomenological approach. *Journal of Medical Ethics*, 47. <https://doi.org/10.1136/medethics-2020-106679>.
- DuBrul, S. A. (2014). The Icarus Project: A counter narrative for psychic diversity. *Journal of Medical Humanities*, 35, 257–271. <https://doi.org/10.1007/s10912-014-9293-5>.
- El Kassar, N. (2022). Epistemische Ungerechtigkeiten in und durch Algorithmen – ein Überblick [Epistemic injustices in algorithms – an overview]. *Zeitschrift für Praktische Philosophie*, 9(1), 279–304. <https://doi.org/10.22613/zfpp/9.1.11>.
- Faissner, M., Juckel, G., & Gather, J. (2022). Testimoniale Ungerechtigkeit gegenüber Menschen mit psychischer Erkrankung in der Gesundheitsversorgung. Eine konzeptionelle und ethische Analyse [Testimonial injustice against people with mental disorders in health care. A conceptual and ethical analysis]. *Ethik in der Medizin*, 34(2), 145–160. <https://doi.org/10.1007/s00481-021-00666-7>.
- Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*, 21, e13216. <https://doi.org/10.2196/13216>.
- Ford, J., Thomas, F., Byng, R., & McCabe, R. (2019). Exploring how patients respond to GP recommendations for mental health treatment: An analysis of communication in primary care consultations. *BJGP Open*, 3(4). <https://doi.org/10.3399/bjgpopen19X101670>.
- Fricker, M. (2007). *Epistemic injustice. Power and the ethics of knowing*. Oxford University Press.
- Fried, E. I., Rieble, C., & Proppert, R. K. K. (2023). Building an early warning system for depression: Rationale, objectives, and methods of the WARN-D study. [preprint] <https://doi.org/10.31234/osf.io/9qcvs>.
- Fuchs, T. (2013). Depression, intercorporeality, and interaffectivity. *Journal of Consciousness Studies*, 20(7–8), 219–238.
- Golomb, B. A., McGraw, J. J., Evans, M. A., & Dimsdale, J. E. (2007). Physician response to patient reports of adverse drug effects. *Drug Safety*, 30, 669–675. <https://doi.org/10.2165/00002018-200730080-00003>.
- Hoffman, G. A. (2019). Public mental health without the health? Challenges and contributions from the mad pride and neurodiversity paradigms. In K. Cratsley, & J. H. Radden (Eds.), *Mental health as public health: Interdisciplinary perspectives on the ethics of prevention* (pp. 289–326). Elsevier/Academic.
- Horwitz, A. V. (2012). Social constructions of mental illness. In H. Kincaid (Ed.), *The Oxford handbook of philosophy of social science* (pp. 559–578). Oxford University Press.
- Horwitz, A. V., & Wakefield, J. C. (2012). *The loss of sadness. How psychiatry transformed normal sorrow into depressive disorder*. Oxford University Press.
- Huda, A. S. (2021). The medical model and its application in mental health. *International Review of Psychiatry*, 33, 463–470. <https://doi.org/10.1080/09540261.2020.1845125>.
- Hull, G. (2023). Dirty data labeled dirt cheap: Epistemic injustice in machine learning systems. *Ethics and Information Technology*, 25, 38. <https://doi.org/10.1007/s10676-023-09712-y>.
- Jain, S. H., Powers, B. W., Hawkins, J. B., & Brownstein, J. S. (2015). The digital phenotype. *Nature Biotechnology*, 33, 462–463. <https://doi.org/10.1038/nbt.3223>.
- Kidd, I. J., & Carel, H. (2018). Harms and wrongs in epistemic practice. Healthcare practice, epistemic injustice, and naturalism.

- Royal Institute of Philosophy Supplement*, 84, 211–233. <https://doi.org/10.1017/s1358246118000620>.
- Kidd, I. J., Spencer, L., & Carel, H. (2022). Epistemic injustice in psychiatric research and practice. *Philosophical Psychology*. <https://doi.org/10.1080/09515089.2022.2156333>.
- Kiesler, D. J. (1999). *Beyond the disease model of mental disorders*. Greenwood Publishing Group.
- Kingma, E. (2007). What is it to be healthy? *Analysis*, 67(294), 128–133. <https://doi.org/10.1093/analys/67.2.128>.
- Klugman, C. M. (2021). Black boxes and bias in AI challenge autonomy. *American Journal of Bioethics*, 21, 33–35. <https://doi.org/10.1080/15265161.2021.1926587>.
- Köhnen, M., Dirmaier, J., & Härter, M. (2019). Potenziale und Herausforderungen von E-Mental-Health-Interventionen in der Versorgung psychischer Störungen [Potentials and challenges of e-mental health interventions in mental health care]. *Fortschritte der Neurologie – Psychiatrie*, 87(3), 160164. <https://doi.org/10.1055/a-0853-2568>.
- Kurs, R., & Grinshpoon, A. (2018). Vulnerability of individuals with mental disorders to epistemic injustice in both clinical and social domains. *Ethics & Behavior*, 28, 336–346. <https://doi.org/10.1080/10508422.2017.1365302>.
- Lakeman, R. (2010). Epistemic injustice and the mental health service user. *International Journal of Mental Health Nursing*, 19, 151–153. <https://doi.org/10.1111/j.1447-0349.2010.00680.x>.
- Leaning, I. E., Ikani, N., Savage, H. S., Leow, A., Beckmann, C., Ruhé, H. G., & Marquand, A. F. (2024). From smartphone data to clinically relevant predictions: A systematic review of digital phenotyping methods in depression. *Neuroscience & Biobehavioral Reviews*, 158, 105541. <https://doi.org/10.1016/j.neubiorev.2024.105541>.
- 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, e32. <https://doi.org/10.2196/mental.9423>.
- Mayer, G., Gronewold, N., Alvarez, S., Bruns, B., Hilbel, T., & Schultz, J. H. (2019). Acceptance and expectations of medical experts, students, and patients toward electronic mental health apps: Cross-sectional quantitative and qualitative survey study. *JMIR Mental Health*, 6, e14018. <https://doi.org/10.2196/14018>.
- Miller Tate, A. J. (2019). Contributory injustice in psychiatry. *Journal of Medical Ethics*, 45, 97–100. <https://doi.org/10.1136/medethics-2018-104761>.
- Mitchell, A. J., Vaze, A., & Rao, S. (2009). Clinical diagnosis of depression in primary care: A meta-analysis. *The Lancet*, 374(9690), 609–619. [https://doi.org/10.1016/S0140-6736\(09\)60879-5](https://doi.org/10.1016/S0140-6736(09)60879-5).
- Mitchell, A. J., Rao, S., & Vaze, A. (2011). Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. *Journal of Affective Disorders*, 130(1–2), 26–36. <https://doi.org/10.1016/j.jad.2010.07.028>.
- Moitra, M., Santomauro, D., Collins, P. Y., Vos, T., Whiteford, H., Saxena, S., & Ferrari, A. J. (2022). The global gap in treatment coverage for major depressive disorder in 84 countries from 2000–2019: A systematic review and bayesian meta-regression analysis. *PLoS Medicine*, 19, e1003901. <https://doi.org/10.1371/journal.pmed.1003901>.
- Moshe, I., Terhorst, Y., Asare, O., Sander, K., Ferreira, L., Baumeister, D., Mohr, H., D., & Pulkki-Räback, L. (2021). Predicting symptoms of depression and anxiety using smartphone and wearable data. *Frontiers in Psychiatry*, 28(12), 625247. <https://doi.org/10.3389/fpsy.2021.625247>.
- Müller, V. C. (2021). Ethics of artificial intelligence and robotics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* Retrieved January 18, 2022, from <https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/>.
- Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, 2, 100347. <https://doi.org/10.1016/j.patter.2021.100347>.
- Pohlhaus, G. (2012). Relational knowing and epistemic injustice: Toward a theory of willful hermeneutical ignorance. *Hypatia*, 27(4), 715–735. <https://doi.org/10.1111/j.1527-2001.2011.01222.x>.
- Polhemus, A., Simblett, S., Dawe-Lane, E., Gilpin, G., Elliott, B., Jilka, S., Novak, J., Nica, R. I., Temesi, G., & Wykes, T. (2022). Health tracking via mobile apps for depression self-management: Qualitative content analysis of user reviews. *JMIR Human Factors*, 9, e40133. <https://doi.org/10.2196/40133>.
- Pozzi, G. (2023a). Automated opioid risk scores: A case for machine learning-induced epistemic injustice in healthcare. *Ethics and Information Technology*, 25, 3. <https://doi.org/10.1007/s10676-023-09676-z>.
- Pozzi, G. (2023b). Testimonial injustice in medical machine learning. *Journal of Medical Ethics*, 49, 536–540. <https://doi.org/10.1136/jme-2022-108630>.
- Puddifoot, K. (2019). Stereotyping patients. *Journal of Social Philosophy*, 50(1), 69–90. <https://doi.org/10.1111/josp.12269>.
- Radden, J. H. (2012). Recognition rights, mental health consumers and reconstructive cultural semantics. *Philosophy Ethics and Humanities in Medicine*, 13, 6. <https://doi.org/10.1186/1747-5341-7-6>.
- Rashed, M. A. (2019). In defense of madness: The problem of disability. *Journal of Medicine and Philosophy*, 44, 150–174. <https://doi.org/10.1093/jmp/jhy016>.
- Roberts, L. W. (2013). *Community-based participatory research for improved mental healthcare*. Springer.
- Russo, F., Schliesser, E., & Wagemans, J. (2023). Connecting ethics and epistemology of AI. *AI & Society*. <https://doi.org/10.1007/s00146-022-01617-6>.
- Sakakibara, E. (2023). Epistemic injustice in the therapeutic relationship in psychiatry. *Theoretical Medicine and Bioethics*, 44, 477–502. <https://doi.org/10.1007/s11017-023-09627-1>.
- Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2, 420. <https://doi.org/10.1007/s42979-021-00815-1>.
- Saul, J. (2017). Implicit bias, stereotype threat, and epistemic injustice. In I. J. Kidd, J. Medina, & G. Pohlhaus (Eds.), *The Routledge handbook of epistemic injustice* (pp. 235–242). Routledge.
- Schmietow, B., & Marckmann, G. (2019). Mobile health ethics and the expanding role of autonomy. *Medicine Health Care and Philosophy*, 22, 623–630. <https://doi.org/10.1007/s11019-019-09900-y>.
- Scrutton, A. P. (2017). Epistemic injustice and mental illness. In I. J. Kidd, J. Medina, & G. Pohlhaus (Eds.), *The Routledge handbook of epistemic injustice* (pp. 347–355). Routledge.
- Shefer, G., Henderson, C., Howard, L. M., Murray, J., & Thornicroft, G. (2014). Diagnostic overshadowing and other challenges involved in the diagnostic process of patients with mental illness who present in emergency departments with physical symptoms – A qualitative study. *Plos One*, 9(11), e111682. <https://doi.org/10.1371/journal.pone.0111682>.
- Shorey, S., Chee, C. Y. L., Ng, E. D., Chan, Y. H., Tam, W. W. S., & Chong, Y. S. (2018). Prevalence and incidence of postpartum depression among healthy mothers: A systematic review and meta-analysis. *Journal of Psychiatric Research*, 104, 235–248. <https://doi.org/10.1016/j.jpsychires.2018.08.001>.
- Skorburg, J. A., & Yam, J. (2021). 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>.
- Slack, S. K., & Barclay, L. (2023). First-person disavowals of digital phenotyping and epistemic injustice in psychiatry.

- Medicine Health Care and Philosophy*, 26(4), 605–614. <https://doi.org/10.1007/s11019-023-10174-8>.
- Spinazze, P., Rykov, Y., Bottle, A., & Car, J. (2019). Digital phenotyping for assessment and prediction of mental health outcomes: A scoping review protocol. *British Medical Journal Open*, 9, e032255. <https://doi.org/10.1136/bmjopen-2019-032255>.
- Stanghellini, G., & Leoni, F. (2020). Digital phenotyping: Ethical issues, opportunities, and threats. *Frontiers in Psychiatry*, 11, 473. <https://doi.org/10.3389/fpsy.2020.00473>.
- Su, C., Xu, Z., Pathak, J., & Wang, F. (2020). Deep learning in mental health outcome research: A scoping review. *Translational Psychiatry*, 10, 116. <https://doi.org/10.1038/s41398-020-0780-3>.
- Symons, J., & Alvarado, R. (2022). Epistemic injustice and data science technologies. *Synthese*, 200, 87. <https://doi.org/10.1007/s11229-022-03631-z>.
- Talukdar, S. (2021). Undisclosed probing into decision-making capacity: A dilemma in secondary care. *BMC Medical Ethics*, 22, 100. <https://doi.org/10.1186/s12910-021-00669-5>.
- Tekin, Ş. (2020). 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>.
- Theunissen, M., & Browning, J. (2022). Putting explainable AI in context: Institutional explanations for medical AI. *Ethics and Information Technology*, 24(2), 23. <https://doi.org/10.1007/s10676-022-09649-8>.
- Thomas, E., Benjamin-Thomas, T. E., Sithambaram, A., Shankar, J., & Chen, S. P. (2023). Participatory action research among people with serious mental illness: A scoping review. *Qualitative Health Research*, 34(1–2), 3–19. <https://doi.org/10.1177/10497323231208111>.
- Thornicroft, G., Rose, D., & Kassam, A. (2007). Discrimination in health care against people with mental illness. *International Review of Psychiatry*, 19, 113–122. <https://doi.org/10.1080/09540260701278937>.
- Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*, 20, 318–335. <https://doi.org/10.1002/wps.20883>.
- Torrey, E. F., Simmons, W. W., Hancq, S., E., & Snook, J. (2021). The continuing decline of clinical research on serious mental illnesses at NIMH. *Psychiatric Services*, 72, 1342–1344. <https://doi.org/10.1176/appi.ps.202000739>.
- Uusitalo, S., Tuominen, J., & Arstila, V. (2020). Mapping out the philosophical questions of AI and clinical practice in diagnosing and treating mental disorders. *Journal of Evaluation in Clinical Practice*, 27, 478–484. <https://doi.org/10.1111/jep.13485>.
- van Riel, R. (2016). What is constructionism in psychiatry? From social causes to psychiatric classification. *Frontiers in Psychiatry*, 7, 57. <https://doi.org/10.3389/fpsy.2016.00057>.
- von Kardoff, E. (2017). Diskriminierung seelisch Beeinträchtiger [Discrimination of people with mental impairment.] In A. Scherr, A. El-Mafaalani, & E. G. Yüksel (Eds.), *Handbuch Diskriminierung* (pp. 565–592). Springer VS.
- Wardrope, A., & Reuber, M. (2022). The hermeneutics of symptoms. *Medicine Health Care and Philosophy*, 25, 395–412. <https://doi.org/10.1007/s11019-022-10086-z>.
- Wyatt, J. C. (2018). How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? *BMC Medicine*, 16, 225. <https://doi.org/10.1186/s12916-018-1211-7>.
- Young, A., Choi, A., Cannedy, S., Hoffmann, L., Levine, L., Liang, L. J., Medich, M., Oberman, R., & Olmos-Ochoa, T. (2022). Passive mobile self-tracking of mental health by veterans with serious mental illness: Protocol for a user-centered design and prospective cohort study. *JMIR Research Protocols*, 11(8), e39010. <https://doi.org/10.2196/39010>.
- Zarsky, T. (2016). The trouble with algorithmic decisions. *Science Technology & Human Values*, 41, 118–132. <https://doi.org/10.1177/0162243915605575>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.