# Emotion AI Use in U.S. Mental Healthcare: Potentially Unjust and Techno-Solutionist

KAT ROEMMICH SHANLEY CORVITE CASSIDY PYLE NADIA KARIZAT and NAZANIN ANDALIBI University of Michigan, USA

Emotion AI, or AI that claims to infer emotional states from various data sources, is increasingly deployed in myriad contexts, including mental healthcare. While emotion AI is celebrated for its potential to improve care and diagnosis, we know little about the perceptions of data subjects most directly impacted by its integration into mental healthcare. In this paper, we qualitatively analyzed U.S. adults' open-ended survey responses ($n$ = 395) to examine their perceptions of emotion AI use in mental healthcare and its potential impacts on them as data subjects. We identify various perceived impacts of emotion AI use in mental healthcare concerning 1) mental healthcare provisions; 2) data subjects' voices; 3) monitoring data subjects for potential harm; and 4) involved parties' understandings and uses of mental health inferences. Participants' remarks highlight ways emotion AI could address existing challenges data subjects may face by 1) improving mental healthcare assessments, diagnoses, and treatments; 2) facilitating data subjects' mental health information disclosures; 3) identifying potential data subject self-harm or harm posed to others; and 4) increasing involved parties' understanding of mental health. However, participants also described their perceptions of potential negative impacts of emotion AI use on data subjects such as 1) increasing inaccurate and biased assessments, diagnoses, and treatments; 2) reducing or removing data subjects' voices and interactions with providers in mental healthcare processes; 3) inaccurately identifying potential data subject self-harm or harm posed to others with negative implications for wellbeing; and 4) involved parties misusing emotion AI inferences with consequences to (quality) mental healthcare access and data subjects' privacy. We discuss how our findings suggest that emotion AI use in mental healthcare is an insufficient techno-solution that may exacerbate various mental healthcare challenges with implications for potential distributive, procedural, and interactional injustices and potentially disparate impacts on marginalized groups.

CCS Concepts: • **Human-centered computing** → Empirical studies in HCI.

Additional Key Words and Phrases: emotion artificial intelligence, emotion AI, emotion recognition, health care, healthcare providers, mental health, data subjects, justice

**ACM Reference Format:**
Kat Roemmich, Shanley Corvite, Cassidy Pyle, Nadia Karizat, and, Nazanin Andalibi. 2024. Emotion AI Use in U.S. Mental Healthcare: Potentially Unjust and Techno-Solutionist. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 47 (April 2024), 46 pages. https://doi.org/10.1145/3637324

## 1 INTRODUCTION

The United States mental healthcare system is facing growing crises including providers unable to meet demand for mental healthcare services [88, 137, 170] and mental health resources remaining inadequate and unaffordable [36, 74]. These crises disparately impact communities marginalized along dimensions of gender, race, and mental health status [96, 154, 157, 176, 185, 195, 200]. Meanwhile, patients, particularly those with marginalized identities, may experience prejudice and

Authors' address: Kat Roemmich, roemmich@umich.edu; Shanley Corvite, scorvite@umich.edu; Cassidy Pyle, cpyle@umich.edu; Nadia Karizat, nkarizat@umich.edu;
Nazanin Andalibi, andalibi@umich.edu, University of Michigan, Ann Arbor, MI, USA.


ACM 2573-0142/2024/4-ART47
https://doi.org/10.1145/3637324

bias in their interactions with providers and the healthcare system writ large [155], and medical gaslighting, where providers dismiss patients' concerns and privilege biomedical expertise over patients' lived experiences [7]. Even in times of crisis, patients often endure long wait times for diagnoses and treatments [4]. Diagnoses themselves are often inaccurate and embedded with (mis)conceptions of race, class, and gender, impacting subsequent treatment plans [9, 155]. For example, Black children are more likely than white children to be diagnosed with Oppositional Defiant Disorder (ODD) than Attention-Deficit Hyperactive Disorder (ADHD), leading to disparate treatment plans and outcomes [9, 155]. Attempts to improve the state of mental healthcare include the design, development, and deployment of nascent technologies, including e-health applications or "digital therapeutics" [53], teletherapy platforms [60, 78, 128], chatbots [110], virtual reality applications [110], and video games [110].

Among these mental health technologies are applications enabled by emotion artificial intelligence (henceforth referred to as "emotion AI"), a class of technology that attempts to automatically infer human emotion and interact accordingly [91, 101, 133]. Emotion AI applies a range of affective computing and artificial intelligence techniques to various data inputs (e.g., facial expressions, voice, gait, text, biosignals, online behavior) [120, 121, 165] to "sense, learn about, and interact with human emotional life" [121]. While the emotion AI research landscape includes advanced technologies that artificially simulate machine emotions [166], much of the commercially-available emotion AI-enabled mental healthcare applications often leverage "simplistic data science rather than sophisticated AI" [100] to infer and respond to human emotion.

The application of emotion AI to healthcare is purported to benefit the domain by increasing access to care, [168], improving diagnostic tools [103], and providing faster diagnoses through augmented or automated decision-making processes [100]. Examples of emotion AI in mental healthcare abound, and the market for emotion AI technologies continues to grow [124]. For example, the software Ellipsis offers integrations to healthcare providers that use deep learning to analyze patient speech patterns and bio-signals to detect and monitor depression, anxiety, and signs of emotional distress [110]. Similarly, Twill offers an app to healthcare providers that monitors patients' emotional states to personalize its wellbeing interventions [103]. Other emotion AI-enabled mental health technologies include wellbeing interventions with more specific therapeutic targets such as harm prevention [157].

Alongside its promise to improve care, integrating emotion AI into mental healthcare carries risks of adverse consequences to those subjected to the technology. Prior work investigating emotion AI in other contexts (e.g., social media, education, workplace) has identified potential harms associated with its use including individual and societal risks [5], discrimination against marginalized groups [41, 157], privacy intrusion [5, 19, 159, 190], and inadequate and unethical oversight [29, 30].

While potential harms may follow emotion AI use in mental healthcare, its development, design, and implementation is generally driven and promoted by privileged parties [6, 30, 31, 123, 193] (i.e., clinical practice decision-makers and technologists [100]) – largely excluding the perspectives and participation of the relevant social groups (i.e., patients as data subjects) targeted by the technology and thus most vulnerable to its impacts [17, 96, 154, 157, 176, 185, 200]. Yet, a relational ethics [17, 41] perspective positions patients as best equipped to recognize potential harms associated with emotion AI in healthcare, which may otherwise go unnoticed by decision-makers and technologists in more privileged positions. Patients' perspectives can illuminate if and what impacts patients might anticipate from emotion AI use in mental healthcare, and whether proposed emotion AI uses [52, 72, 100, 119, 157, 164, 179] are considered beneficial to patients *themselves*. Ultimately, understanding patients' perspectives can inform future emotion AI development and deployment, or alternatives to this technology, within the mental healthcare domain. The present study centers

the voices of patients as the data subjects[1] who would be directly vulnerable to impacts of emotion AI in mental healthcare to address the following research question:

How do data subjects perceive emotion AI to impact their mental healthcare?

As part of a larger study, we surveyed U.S. adults and analyzed a subset of responses ($n$ = 395), partly representative of the U.S. population and oversampled for marginalized groups (i.e., people of color, gender minorities, and people with current or past-lived experience with mental illness) who may be at more risk of harm and disparate treatment as a result of emotion AI use in healthcare [123, 154, 157, 176, 185, 200]. Participants first answered factorial vignette-based questions derived from various purported uses of emotion AI and other related automatic emotion recognition systems in healthcare [52, 72, 100, 119, 157, 164, 179]. After considering their comfort levels to these scenarios, participants then answered open-ended questions regarding what benefits, harms, undesired impacts, or concerns, if any, they perceived to surface from emotion AI use in healthcare broadly. Our study design for these open-ended questions was intentionally broad to allow participants to conceptualize the impacts most meaningful to *them*. Participants consistently responded with perceived impacts of emotion AI to *mental* healthcare specifically. This paper uses data collected via these open-ended questions to address our research question.

Informed by potential emotion AI uses suggested in vignettes, participants acknowledged how emotion AI could impact data subjects when used to address existing mental healthcare-related issues, including 1) improving mental healthcare assessments, diagnoses, and treatments; 2) facilitating data subjects' mental health information disclosures; 3) identifying potential data subject self-harm or harm posed to others; and 4) increasing involved parties' understanding of mental health. Yet, participants anticipated such applications to negatively impact data subjects' mental healthcare by 1) increasing inaccurate assessments, diagnoses, and treatments along with providers' biases; 2) reducing or removing data subjects' voices and interactions with providers in mental healthcare processes; 3) inaccurately identifying potential data subject self-harm or harm posed to others with implications for negative wellbeing effects; and 4) involved parties misusing emotion AI inferences with negative consequences to (quality) mental healthcare access and privacy.

We situate these findings that our speculative methods surfaced regarding the manifold implications of emotion AI use in mental healthcare (e.g., [5, 41, 157–159, 173, 190]) to argue that emotion AI development and integration in mental healthcare is a proposed *techno-solution* [67, 109] to the various challenges persisting within this high-stakes domain and that is insensitive to values of the data subjects directly impacted by the technology. Techno-solutionism relies on the assumption that technological interventions can be used to solve complex social problems and that such techno-solutions will be advantageous to those involved [67, 109]. Yet, this assumption overlooks the potentially adverse effects of implementing technological interventions and the harms that may be presented to those subjected to the technology. Thus, we end by discussing 1) how proposed emotion AI uses to address present mental healthcare challenges, shrouded by illusions of technological advancement, are insufficient techno-solutions that may exacerbate the very problems emotion AI is implemented to solve; and 2) how emotion AI as a techno-solution can further entrench existing injustices in mental healthcare.

## 2 RELATED WORK

### 2.1 Emotion AI in mental) healthcare: promises, challenges, and ethical concerns

The development and application of emotion AI technologies aiming to identify and/or act on human emotion and other affective phenomena is growing [120, 121, 124]. Emotion AI emerged

[1]We use the term "data subjects" to refer to individuals whose data enables and is processed by emotion AI, and are consequently impacted by emotion AI in practice, in line with HCI scholars' use of the term [41, 73, 157].

from the field of Affective Computing, pioneered by Rosalind W. Picard in 1995 to enhance human interaction with technologies that can predict and/or interact with human emotions [145]. Algorithmic emotion classifications have theoretical underpinnings that invoke historical and ongoing debate regarding the nature of emotion, including Charles Darwin's foundational study of emotions as evolved biological processes [48]; the subsequent James-Lange Theory of Emotions which assumes emotions are responses to physiological changes [86, 104]; and Paul Ekman's Basic Emotions Theory which contends humans universally express a set of basic emotion families physiologically according to six discrete, mutually exclusive categories: anger, disgust, fear, joy, sadness, and surprise [57, 58]. Together, this line of emotion theories has culminated in a "common view" [10] that understands human emotion as biologically determined, universally expressed, and discretely categorized. A large body of work empirically contests said "common view" [10] with an understanding of emotions as constructed by individuals situated in social contexts rather than biologically determined [11], variable across individual and socio-cultural differences rather than universally expressed [44], and structured from multiple overlapping, more fundamental affective dimensions rather than by discrete categorizations [161]. These debates are at the core of current ethical controversy surrounding the scientific foundations of emotion AI [43, 173]. Particularly, the discrete emotion classification schemes that make emotions easily tractable via categorized emotion inferences underlie the majority of emotion AI applications [173] despite their limited reliability, generalizability, specificity, and validity [10] – including in the high-stakes context of mental healthcare [101, 114].

While investigations of *emotion* AI in mental healthcare are nascent [18, 149, 175], existing concerns about broad AI's deployment to clinical practice [95] can be applied to understand the implications of emotion AI's emerging development and implementation in this domain. Primary concerns of broad AI in healthcare include breaches of trust and privacy and lack of transparency in the context of clinical decision support systems [189], clinical chatbots [32], and the broader healthcare system [12, 28, 105, 107]. Moreover, research suggests that the application of AI to clinical practice may adversely affect patient adherence to providers' medical advice [32], the patient-provider relationship [32, 105, 135, 136, 169, 189], and existing clinical workflows [62, 129, 169, 189]. Research investigating AI in clinical settings primarily focuses on technical diagnostic accuracy [95, 177] and bias [3, 28, 95, 136, 141] without directly linking these evaluations to patient benefits, outcomes, or perceptions. Patients' voices and interactions with providers are important aspects of healthcare [21, 55], yet patients' perspectives about incorporating automated technology into the care they receive is largely absent from these assessments.

Emotion AI may implicate unique aspects of and have magnified impacts to mental healthcare as it aims to infer *emotions* – a deeply personal, sensitive, and intimate aspect of individual wellbeing. While the known concerns of emotion AI in healthcare mirror those of AI broadly including patient trust, privacy, and safety [101, 114, 121, 154, 173], recent scholarship investigating patient perceptions of specific use cases of emotion AI-enabled healthcare technologies has expanded the scope of acknowledged concerns to include risks of therapeutic chatbots imposing adverse clinical and psychological effects [101] and mental health risk predictions impairing patient agency [114]. For instance, scholars have explored individuals' attitudes towards being made aware of their mental health risk prediction results [114] and young peoples' experiences using chatbots for mental health-related support [101]. Yet, it remains unclear whether and how these risks may manifest across the wide range of potential emotion AI-enabled mental health applications and are entangled with the potential benefits each may promise. To further understand the potential impacts of this increasingly pervasive technology to mental healthcare, our study centers the perceptions of the data subjects who would be most impacted by the technology [17, 41, 123, 154, 157, 176, 185]

building on existing HCI scholarship that examined other contexts such as education [190], the workplace [41], and social media [157].

### 2.2 Historical legacy of challenges in healthcare and disability contexts

Understanding health-related social movements that have occurred as a result of healthcare challenges experienced by patients, providers, and other relevant involved parties can provide insights into how emotion AI may exacerbate the very challenges it is proposed to solve [142]. Health Social Movements (HSM) entail the mobilization of formal and informal networks (e.g., community organizations, media outlets) around healthcare-related issues [23]. HSMs in the U.S. (e.g., disability rights, patients' rights, women's health) describe, raise awareness around, and resist challenges in healthcare for diverse groups, with shared tensions around equal access to quality care, identity-based discrimination, and the right to self-determination [1, 16, 23, 46, 61, 80, 85, 139, 140, 156, 192]. While detailing the entire history of HSMs is beyond the scope of this paper, we aim to connect the challenges faced by social groups working to be fully recognized as complex human beings worthy of dignity, respect, and autonomy over their lives to those of the data subjects who are directly impacted by emotion AI's integration to mental healthcare, yet are largely excluded from adoption decisions and lack the social power to contest its use.

HSMs recognize patient-healthcare provider interactions as sites of challenges related to unequal power dynamics, discriminatory language and treatment, and unfounded expectations based on identity or perceived mental capacity. For example, in the mid-1960s, individuals in women's health movements were concerned with sexist interactions with medical professionals that impacted their care delivery and access to health information, citing shared experiences with doctors withholding their health information and dismissing their needs and concerns when making medical decisions [156]. The women's health movement raised alarms around uninformed consent in the medical decision-making process and violations of patients' bodily autonomy [156]. The 1970s saw the psychiatric survivor/ex-patient movement raise awareness about involuntary institutionalization and medical treatments (e.g., electric shock therapy) [139, 152], especially for those marginalized along dimensions of race, sexual orientation, socioeconomic status, and non-conformity to social norms [1], highlighting clinicians' coercion of vulnerable people to comply with invasive medical interventions that were inappropriate for most patients [1]. More recently, citizen-science alliances have forged wherein everyday people collaborate on various, historically exclusive processes (e.g., research) alongside medical authorities and scientific communities [16, 22]. For example, AIDS activists in the 1980s and 90s organized to ensure that their personal experience would shape future AIDS-related scientific research, as opposed to remaining ignored and perceived to lack epistemic authority [59, 85]. Today's environmental justice and women's health efforts challenge inequitable health outcomes and the disparate attention paid by researchers to different communities' health needs [16, 22]. Years of disability rights activism throughout the 20th century similarly challenged disabled persons' exclusion from equal access to healthcare due to inaccessible physical barriers [131, 138, 140, 182], culminating with the passing of the Americans with Disabilities Act (ADA) in 1990 [131, 139, 152] that requires medical providers to provide "full and equal access to their healthcare services and facilities" and make accommodations so anyone may receive access to care, regardless of ability status [182].

Altogether, health-related social movements show how individuals and communities have organized against the inequitable distribution of resources in healthcare contexts based on one's identity, health, or perceived abilities, and remind us that patients' historical exclusion as authority figures in health decision-making [156] intersects with contours of unequal power dynamics shaped by social position and compounded by identity dimensions including race, class, and ability. This study is situated in and motivated by these longer histories of healthcare-related social justice movements

that seek to center the perceptions of those made most vulnerable in seeking and receiving care by privileging patients' lived experiences, valuing their complex humanity, and recognizing their right to societal participation, respect, and self-determination as we center the voices of those who stand to be most impacted by the deployment of emotion AI in mental healthcare.

## 3 METHODS

This section outlines the survey design, recruitment and participation, data analysis, limitations, and opportunities for future work.

### 3.1 Study design

*3.1.1 Survey design.* The survey, developed as part of a larger research program, included a total of 56 factorial vignettes organized by two sets of contexts (employment and healthcare), with 28 vignettes specific to healthcare. Following vignettes, participants answered open-ended questions to share their perceptions of benefits and harms they associated with emotion AI in healthcare – this paper's focus. Participants were also invited to reflect upon how/if aspects of their identity shaped their responses to survey questions, analysis of which was beyond the scope of this paper. While this paper does not analyze participants' responses to vignettes, we provide details about the survey design as a whole for transparency and context.

The vignettes asked participants to rate their comfort with their provider using a program to automatically detect their emotional state from either 1) "records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it)" or 2) "images or video of what you look like, based on your facial expressions recorded from your daily activities and device use," for 14 different purposes (described in Section 3.1.2 and included in Appendix A). To ensure a common understanding of "emotional state," we included the following definition at the top of the survey: "Emotional state refers to your emotions and moods, including but not limited to stress, anxiety, depression, boredom, calmness, fear, fatigue, attentiveness, happiness, sadness, disgust, surprise, and/or anger." We used this broad definition based on existing emotion classification schemes [19]. Participants indicated their comfort for each vignette.

After completing survey responses to the vignettes, participants were asked to answer four open-ended questions (available in Appendix B). Participants provided their perceptions by answering what, if any, 1) benefits they anticipate from emotion AI in healthcare, 2) harms or other undesired impacts they anticipate from emotion AI in healthcare, 3) other concerns they have about emotion AI, and 4) aspects of their identity (broadly construed) may have shaped their (dis)comfort with emotion AI in healthcare. We ran a pilot survey before data collection to test the survey design and to determine if any changes were necessary (e.g., concerning respondent fatigue and clarity of survey questions). We concluded that no additional changes were necessary and that the survey was ready for data collection.

This paper contributes findings limited to our analysis of participants' responses to the first three open-ended questions regarding potential impacts (i.e., benefits, harms) participants anticipated with emotion AI in healthcare.

*3.1.2 Factorial vignettes.* The vignettes participants responded to before answering the open-ended questions and that we analyze in this paper varied by 14 purposes for which emotion AI may be used in healthcare, based on the following emotion AI use cases proposed in prior work: facilitating early detection of mental and neurological illnesses [52, 72], providing timely and potentially more accurate mental health and wellbeing insights and interventions (in comparison to a human provider) [100, 119, 157, 179], equipping healthcare providers with increased understanding of

patients at individual and group levels [31, 100], detecting or preventing potential self-harm or harm to others [52, 157, 164, 179], and supporting academic research [100, 157].

As vignettes were presented to participants before answering the open-ended questions, their responses to the open-ended questions may have primed. Prior work suggests that because people tend to underestimate the potential benefits and risks of emerging technology, initial priming is an appropriate methodological strategy to elucidate participants' perceived impacts by providing context and knowledge that, if not presented, may otherwise lead to uninformed and uncertain responses [68, 70, 77, 181, 197]. To account for potential negative priming associated with negative public perceptions of emotion AI [5], we did not explicitly name "emotion AI" in the vignettes. Because people have varying degrees of familiarity with technology, it was important to ensure that the definition provided was clear, concrete, descriptive, and to the extent possible, neutral. With these considerations, we were still able to draw perceptions of *emotion AI*, as the vignettes referred to a software or "intelligent computer program" that uses common emotion AI data inputs [121]. Essentially, as participants responded to their perceptions of the described intelligent software, they shared their perceptions on being subjected to emotion AI in healthcare.

We designed factorial vignettes based on best practices [15, 87, 111, 115, 116, 188]. Factorial vignettes are useful to bring forth perceptions and attitudes about phenomena [87]. Thus, the vignettes included in this study were designed to be speculative as participants were asked to imagine if their healthcare provider(s) used emotion AI as described in vignettes, a widespread approach in vignette designs [15, 116]. Speculation is a powerful tool to investigate people's perceptions of technologies' ethical implications [63] especially when they may not have direct experience with said technology, and how technologies can impact who uses and interacts with them [199]. With regard to emerging technology such as emotion AI, speculation through vignettes can surface perceptions and values toward the technology, though people may not be aware of being subjected to it [64]. While speculating on participants' perspectives of emotion AI use in the healthcare context may not directly map to how people would behave in practice, prior work suggests that people would act similarly to how they imagine they would [84, 151]. Though the results of the vignette portion are out of the scope of this paper and may limit the nuance and depth of participants' voices in their open-ended responses, our use of factorial vignettes to surface peoples' perspectives about being subjected to emotion AI in healthcare provides valuable insights into what their attitudes towards emotion AI use in healthcare may be in practice.

### 3.2 Recruitment and participation

This paper analyzes a dataset of 395 U.S. adults' open-ended responses to a survey examining perceptions of emotion AI recruited through Prolific, an online survey recruitment platform. We collected a total of 755 responses and analyzed a subset ($n$ = 395) informing this paper's findings.

We developed the subset of 395 responses by merging a U.S. representative sample ($n$ = 296) (on the axes of sex, age, and race/ethnicity) and an oversample of participants who identified with at least one marginalized identity ($n$ = 99) which we define for the purposes of this paper as individuals who identify as a person of color, gender minority (i.e., transgender, non-binary), and/or individual with current or past-lived experience(s) with mental illness(es) – groups that may disparately experience harms from emotion AI [96, 154, 157, 176, 185, 195, 200]. Our U.S. representative sample was collected using Prolific's representative sample feature that automatically selects participants based on age, sex, and race/ethnicity and is representative of the national population. Then, we ran separate recruitment efforts using Prolific's pre-screened criteria to oversample for participants who identified as a person of color, gender minority, and/or individual with current or past-lived experiences with mental illness(es) ($n$ = 455). From the oversampled dataset, we randomly subsampled participants ($n$ = 109) to include those who fully completed

the survey *and* identified with at least one marginalized identity. We then combined the random subsample from the oversample with the representative sample ($n$ = 409). After merging, we removed 14 participants for disingenuous, blank, and duplicate responses to the open-ended questions, resulting in a total of 395 analyzed responses. For duplicate responses, we included only the participants' first submission in the analyzed dataset.

We decided to develop a dataset that included a U.S. representative sample *and* an oversample of individuals with at least one marginalized identity to highlight the ways marginalized individuals may experience more harm in seeking mental healthcare, and may perceive and experience more harm stemming from the use of emotion AI broadly [123, 154, 157, 176, 185, 200]. It is well documented that marginalized populations – including LGBTQ+ folks [24, 82, 171, 194], disabled and mentally ill folks [39, 40, 180], people of color [24, 25, 92, 102], among other groups – face barriers and stigmatization when seeking mental healthcare. More specifically, these groups experience similar barriers of service denial [102, 194], affordability [171], and stigmatizing or microaggressive interactions with service providers [25, 39, 40, 92, 171]. The particularities of these stigmatizing or microaggressive interactions, importantly, likely manifest differently depending on the marginalized identity a patient embodies, but these high-level problems have been demonstrated across marginalized identity groups. For instance, according to the American Journal of Managed Care, among those considered "vulnerable populations" in healthcare are racial or ethnic minorities and those with chronic conditions (e.g., mental illness) [176]. Additionally, prior work highlights how emotion AI may disproportionately negatively impact those with mental and physical disabilities and racial and ethnic minorities [123, 153]. Importantly, these negative impacts may be experienced differently between marginalized groups. For example, folks with mental illness may experience greater stigma and decreased work opportunities as a result of the uptake of emotion AI that may discriminate against those with mental illness in hiring and the workplace [123]. Meanwhile, other work shows that Black faces are interpreted by emotion recognition algorithms as being more angry and contemptuous than their white counterparts, with downstream implications for exacerbating stereotypes and differential treatment in the commercial application of these technologies [153]. Additionally, emerging scholarship notes the distinct and/or exacerbated harms emotion AI may confer for those with marginalized identities. For instance, Monteith et al. highlight the heightened concerns of emotion AI use for people with mental illness and disabilities, as the technology may lead to increased stigma around mental illness and disability and may create norms and presumptions about people with mental illness and disability from datasets that exclude these identities [123]. Aligned with prior work [76], oversampling for these groups allowed us to surface potential participant responses that may not have surfaced otherwise, including in a nationally representative sample. A full breakdown of the sample is available in Appendix C, which includes a breakdown of the representative sample and oversample, and Appendix D, which includes aggregate counts of participants' demographics.

We note that during our data analysis, we reached saturation, and no new themes surfaced; therefore, we did not increase the sample size. Moreover, although the oversample of marginalized participants is only a fraction of the entire sample included in this study, a large majority (85% or 336) of the participants included in our analysis hold at least one marginalized identity (on the axes of race/ethnicity, gender, and past/current experience with mental illness). We provide further detail of participant demographics in Appendix D. Participants were compensated with $3.80 per Prolific's suggested calculation. The average survey completion time was 24 minutes.

### 3.3 Analysis

We conducted an iterative qualitative analysis of participants' answers to three open-ended questions described in Section 3.1.1.

| Codebook themes | Alpha binary |
|---|---|
| Perceived potential benefits of emotion AI in healthcare | 0.881 |
| Perceived potential concerns of emotion AI in healthcare | 0.837 |
| Average alpha binary across relevant themes | 0.859 |

Table 1. Alpha binaries and average of alpha binaries of codebook themes relevant to this study

We developed a codebook through several coding exercises to create a common understanding among the final two coders involved in data analysis. The first coding exercise was done on 50 randomly selected open-ended responses from the dataset of 395 participants by five coders trained in qualitative coding. The five coders independently open-coded the random subset of 50 responses. They then met to discuss their codes, observe overall themes, and agree upon codes to include in the codebook. Based on this meeting, the last author then developed an initial codebook by categorizing open codes into parent codes. In a second coding exercise, three coders worked together on the initial codebook and applied the revised codebook to a subset of another 35 randomly selected responses. Next, the three coders met with the last author to discuss their codes, resolve disagreements, and develop a finalized codebook.

Once the codebook was finalized, the final two coders conducted additional coding exercises to establish inter-rater reliability (IRR) [118] using ATLAS.ti's calculation for Krippendorff's alpha binary. These two coders reached an IRR score above .75 ("acceptable" [118] for IRR), as seen in Table 1, after two rounds of coding random subsets of 20 responses using the final codebook. To identify and resolve disagreements in codes after the first round, the two coders met to discuss their disagreements and rationales to reach a consensus, ensuring consistent application of all codes.

The same two coders divided and coded the remaining data after establishing IRR. If new themes emerged in the remaining data, the two coders could incorporate new codes upon discussion to ensure consistency. Though no new codes surfaced in this process, this allowed our analysis to remain flexible and to capture all possible themes to the best of our ability. Upon completing coding, the whole research team reconvened to identify and discuss the themes regarding participants' perceived potential benefits and harms associated with emotion AI in healthcare. Major themes surrounding perceived potential emotion AI impacts emerged that relate to present challenges in *mental* healthcare, which we focus on in this study. Participants interpreted and perceived the open-ended questions as predominantly relevant to the *mental* healthcare context, likely because emotion AI technology aims to infer upon data subjects' mental and emotional health. We note that our methodological choice to qualitatively analyze a representative sample and oversample was not to provide quantitative, statistical results but to illustrate insightful recurring themes that emerged in our analysis.

### 3.4 Limitations and opportunities

This study's limitations are three-fold: First, we acknowledge the potential framing bias that may have emerged from the survey vignettes (presented to participants *before* answering the open-ended questions) that incorporated purported emotion AI uses. The vignettes were constructed as neutrally as possible. However, we acknowledge that suggested potential emotion AI uses may still have influenced participants to view emotion AI positively due to framing bias [2, 8, 144, 186, 187]. On the other hand, the vignettes did not indicate any potential harms, and so may have avoided negatively influencing participants' perspectives described in their perceived concerns. However, we recognize the possibility that vignettes may have still elicited immediate negative reactions toward emotion AI use, leading to potentially negative perceptions that could shape their responses to the open-ended questions presented afterward. Nonetheless, we note that the open-ended questions analyzed in

this paper were worded flexibly to allow participants to develop their own perceptions of both perceived benefits and harms of emotion AI use by their healthcare provider(s), which may have mitigated potential influence from the vignettes. However, as we note in Section 3.1.1, priming participants can be useful for understanding people's perceptions of emerging technology as people often underestimate its potential impacts [68, 70, 77, 181, 197]. Furthermore, demographic questions that could have influenced participants' responses were included at the end of the survey to avoid the potential influence of these questions on participants' responses.

Second, while representative samples can be powerful, we supplemented our representative sample with an oversample of marginalized groups to ensure that we do not ignore the perspectives of those who may be most impacted by emotion AI use by their healthcare provider(s). Though all participants were asked to rate their comfort level from the position of a patient in the healthcare context in which emotion AI is used, we note potential limitations regarding our dataset and methodological choices that were constrained by Prolific's (and other services) available pre-screened participant pools, which were automatically selected based on a distribution of age, sex, and race/ethnicity representative of the U.S. national population. To scope the survey, we did not collect *all* data that may be relevant to participants' perceptions of emotion AI use (e.g., technological literacy, job title, experience with emotion AI). Thus, our analyzed dataset may not be representative of and generalizable to the different subpopulations for which we oversampled or to other contexts in which emotion AI may be implemented (e.g., law enforcement, social media, education) as people's perceptions about technology, data sharing, and emotional expression are contextually situated [94, 119, 130]. Yet, through oversampling, we surfaced insights from these subpopulations that could be built on in future work. For instance, future work could conduct participatory design sessions with those living with mental illness(es) to envision technological designs that meet their identified and desired needs while attending to relevant concerns identified in this study or take a quantitative approach to examine and compare the perspectives of emotion AI data subjects across different contexts and identities to assess the prevalence of our qualitative analysis findings regarding marginalized groups on a larger scale.

Lastly, we acknowledge the potential limitations of our decision to analyze the responses to three open-ended questions to effectively investigate patient perceptions of emotion AI. We note that, initially, we were unsure we would gain such rich data from the open-ended questions, and we were thrilled that we did. Indeed, Braun et al. note that "qualitative surveys are an exciting, flexible method with numerous applications, and advantages for researchers and participants alike," and suffer from "(misplaced) assumptions about qualitative survey data lacking depth" [20]. This approach has also made its way to prior CSCW/HCI scholarship (e.g., [75, 76]), and we encourage future work to use similar approaches, especially in contexts concerning emerging technologies. This approach has the potential for deep, generative, and scaled understanding of phenomena [20], which can be supplemented with other methods in the future, drawing from insights generated. For example, future work could potentially build on these findings using other methods, including interviews and/or walkthroughs with participants who have encountered emotion AI technology used in mental healthcare and other contexts to gain insight from participants with first-hand experience. That said, it is possible that future work with other samples or with different study designs may uncover different perceptions. Additionally, we acknowledge that including data from the survey's vignette portion would have provided other insights; however, coupling the two could have led to losing nuance and depth that may, in effect, suppress the voices we aim to magnify in this paper. Nonetheless, our findings provide valuable and timely insights regarding data subjects' perceptions of potential emotion AI impacts to mental healthcare while ensuring that we highlight those who embodied at least one marginalized identity that we focus on in this work.

| **Mental healthcare provisions** |
|---|
| *Perceived potential impacts of emotion AI use when addressing challenges in mental healthcare:*<br>(1) Improve mental healthcare assessments, diagnoses, and treatments |
| *Perceived potential negative implications of using emotion AI in mental healthcare:*<br>(1) Increase inaccurate assessments, diagnoses, and treatment<br>(2) Increase providers' biases |
| **Data subjects' voices** |
| *Perceived potential impacts of emotion AI use when addressing challenges in mental healthcare:*<br>(1) Facilitate data subjects' mental health information disclosures |
| *Perceived potential negative implications of using emotion AI in mental healthcare:*<br>(1) Reduce or remove data subjects' voices and interactions with providers |
| **Monitoring data subjects for potential harm** |
| *Perceived potential impacts of emotion AI use when addressing challenges in mental healthcare:*<br>(1) Identify potential data subject self-harm or harm posed to others |
| *Perceived potential negative implications of using emotion AI in mental healthcare:*<br>(1) Inaccurate identifications of self-harm or harm posed to others<br>(2) Negative well-being effects |
| **Involved parties' understanding and uses of mental health inferences** |
| *Perceived potential impacts of emotion AI use when addressing challenges in mental healthcare:*<br>(1) Enhance involved parties' understanding of mental health |
| *Perceived potential negative implications of using emotion AI in mental healthcare:*<br>(1) Creating barriers to accessing (quality) mental healthcare<br>(2) Intruding on data subjects' privacy |

Table 2. Breakdown of perceived emotion AI impacts to mental healthcare data subjects

## 4 FINDINGS

Our findings illustrate participants' perceptions of potential impacts that emotion AI in mental healthcare could have on data subjects in practice, surfaced from our analysis of their open-ended responses to survey questions regarding the benefits and harms/concerns they anticipate with integrating the technology in healthcare.

We identified four main perceived impacts emotion AI may pose to data subjects when used to address the following existing mental healthcare challenges: 1) improve mental healthcare assessments, diagnoses, and treatments; 2) facilitate data subjects' mental health information disclosures; 3) identify potential data subject self-harm or harm posed to others; and 4) increase involved parties' understanding of mental health. While participants shared perceptions that emotion AI may be beneficial, they also raised concerns regarding how the technology may, consequently, exacerbate extant challenges in mental healthcare and harm emotion AI data subjects: 1) increase inaccurate assessments, diagnoses, and treatments along with providers' biases; 2) reduce or remove data subjects' voices and interactions with providers in mental healthcare processes; 3) inaccurately identify potential data subject self-harm or harm posed to others with implications for negative wellbeing effects; 4) involved parties' misuse of emotion AI inferences with consequences to (quality) mental healthcare access and data subjects' privacy. Table 2 maps each perceived impact of emotion AI with its associated negative implication under the following aspects of mental healthcare: 1) mental healthcare provisions, 1) data subjects' voices, 3) monitoring data subjects for

potential harm, and 4) involved parties' understandings and uses of mental health inferences. We organize the presentation of our results accordingly.

To contextualize our findings, we include 1) overall percentages and counts of participants who shared responses relevant to each finding; 2) percentage of the overall count representing participants who identified with at least one marginalized identity that we oversampled (i.e., a person of color, gender minority, lived experience with mental illness) to foreground participants who may experience exacerbated harms from emotion AI use [17, 123, 154, 157, 176, 185, 200]; and 3) the relevant factorial vignette(s) in footnotes that we interpret as related to each finding.

### 4.1 Mental healthcare provisions

Inaccurate, inefficient, and biased mental healthcare provisions contribute to inadequate healthcare [1, 33, 99, 113, 156, 174, 198]). Some participants recognized emotion AI's potential to mitigate these mental healthcare challenges by enabling accurate, efficient, and unbiased mental health assessments, diagnoses, and treatments.[2] However, many participants also explicated ways that using emotion AI to improve existing mental healthcare provisions may exacerbate already inadequate mental healthcare by producing inaccurate mental health inferences and heightening providers' biases against patients (i.e., emotion AI's data subjects). This section first highlights participants' perceptions of how using emotion AI could improve various mental health provisions, followed by the perceived potential problems that may arise from such uses.

*4.1.1 Perceived potential impacts of emotion AI: improve mental healthcare assessments, diagnoses, and treatments.* Some participants acknowledged how emotion AI may improve mental healthcare provisions by rectifying inaccurate, inefficient, and biased mental health assessments, diagnoses, and treatments carried out by human mental healthcare providers. Providers' failure to meet data subjects' mental health needs can negatively affect their mental and overall health, which participants recognized emotion AI to potentially mitigate by addressing inaccuracies, inefficiencies, and biases within mental healthcare.

**Addressing inaccurate mental healthcare provisions.** 8.4% of participants ($n$ = 33), 84.7% of whom identified with at least one marginalized identity, mentioned that emotion AI may mitigate inaccurately assessed, diagnosed, or treated mental health conditions. For instance, P57, a white man with a mental health condition, described how emotion AI could potentially lead to more accurate mental health diagnoses by detecting patterns that (often inattentive) human providers may miss: "*machines are great at picking up things that humans aren't and vice versa, so a doctor augmenting their diagnoses and treatments with various robots and AI assistants have major potential to improve care across the board.*" P23, a Southeast Asian woman with a mental health condition, echoed P57's sentiments: "*sometimes doctors are busy writing notes or [are] distracted...the system could help detect things the doctor didn't notice.*" P57 and P23's comments point towards extant challenges whereby inattentive or overworked providers may overlook patients' needs, consequently provisioning inaccurate diagnoses and treatments. Integrating emotion AI that may be better at "*picking things up*" than *"busy"* providers is perceived to potentially facilitate more accurate care provisions.

**Facilitating more efficient mental healthcare provisions.** 52.7% of participants ($n$ = 208), 76.4% of whom identified with at least one marginalized identity, acknowledged that emotion AI could address current inefficiencies in detecting and diagnosing mental health conditions. P7,

[2]Acknowledged potential uses of emotion AI to improve mental healthcare provisions map to factorial vignettes that posed the use of emotion AI in healthcare to infer the mental health state of patients (on a group level and individually), infer patients in need of wellbeing support, assess the overall health of individual patients, diagnose mental health illness and neurological disorders in patients earlier than otherwise possible, automatically alerting healthcare provider(s) when patients may need support, and avoiding human judgment and subjectivity.

a transgender, non-binary white person with multiple mental health conditions, described how they were "*diagnosed with ADHD later on in life, so definitely making resources so that people can get diagnoses and properly treated faster would have helped [them].*" For P7, resources that could have aided in an earlier diagnosis and treatment would have benefited them. P33, a white woman who reported no mental health conditions, shared similar thoughts that "*it could be nice for the program to notice a particular health concern before [she] did to facilitate faster treatment.*" These examples highlight how current mental healthcare processes are ill-equipped to provide efficient diagnoses and treatments. This present challenge relates to unique diagnostic difficulties that are partially attributed to a lack of "objective" diagnostic tests in mental healthcare [93]. Participants perceived emotion AI could potentially address these deficiencies by facilitating faster diagnosis and appropriate care to data subjects, reflecting biomedical virtue rhetoric [13, 146], which promotes an uncontested ideal that privileges a "praxis of goodness" within healthcare and legitimates the deployment of new technologies and data practices within its domain (e.g., to facilitate early diagnosis or access to treatment) without adequate critical examination [146].

**Mitigating providers' biases in mental healthcare provisions.** 6.3% of participants ($n$ = 25), 84% of whom identified with at least one marginalized identity, shared perceptions that emotion AI could potentially mitigate the role of mental healthcare providers' biases in care provisioning. Participants acknowledged the possibility that emotion AI may augment providers' decisions with unbiased inferences, reducing the possibility for providers' biases to fully account for mental health assessments, diagnoses, and treatments. For instance, P334, a white transgender man with multiple health conditions, stated, "*I'm an adult ADHD person and would have benefited GREATLY from technology such as this had it been available when I was younger, as the path to my diagnosis was arduous and oftentimes hindered by non-objective professionals.*" P334 described how his difficult experience involving biased providers impeding his timely diagnosis could have been improved with more objective approaches like that promised by emotion AI. Similarly, P95, a Black woman with multiple mental health conditions, stated, "*I figure with all the experiences I've had with human doctors concerning my mental health and physical health... a program that's able to access a great deal of information and give an unbiased evaluation, couldn't be any worse.*" Based on her personal experiences with insufficient care, P95 thought that the "*unbiased evaluation[s]*" emotion AI promises at the very least would not be a worse alternative to biased providers. P95's experiences are situated in a longer history of medical gaslighting and disparate treatments for Black patients that is entangled with majority white providers' subjective biases against Black patients and other minority patients [27, 167]. Overall, the experiences participants shared highlight the present mental healthcare challenge of biased providers negatively affecting if and how patients receive appropriate mental healthcare and their perceptions that emotion AI could potentially mitigate mental healthcare provisioning hindered by providers' biases.

*4.1.2 Perceived potential negative implications of emotion AI: increase inaccurate assessments, diagnoses, and treatments.* Though some participants acknowledged that emotion AI could potentially address persistent challenges of inadequate mental healthcare provisions, many also perceived how emotion AI may instead worsen them. 51.4% of participants ($n$ = 203), 76.8% of whom identified with at least one marginalized identity, noted how emotion AI's potentially inaccurate inferences could worsen already inaccurate assessments, diagnoses, and treatments [9, 99, 113, 174, 198]. P321, a bi-racial woman with multiple mental health conditions, mentioned that "*culturally, expression can vary depending on many factors, which might lead to inaccurate readings,*" depicting doubt that emotion AI could accurately account for complexities in emotional expression across cultures and individual differences. Such concerns are grounded in existing literature [10, 43, 94, 191], and may be heightened for bi-racial/bi-cultural individuals like P321 with mixed identities. P321's concerns

regarding potential algorithmic biases highlight how inaccurate emotion inferences may have harmful ramifications to data subjects' diagnoses and treatments, contesting earlier acknowledgments that inferences, if accurate, would be beneficial as reported in Section 4.1.1. Thus, supplanting care provisioning processes with emotion AI, itself known to demonstrate poor rates of accuracy [10] and perpetuate demographic biases [54, 98], may be an inadequate solution to addressing existing problems of inaccurate assessments, diagnoses, and treatments in mental healthcare, and in effect could *exacerbate* these problems by automatically reproducing them on a large scale.

Participants' concerns regarding potentially inaccurate inferences also illustrate the possibility for providers to take these inferences at face valueand subsequently dismiss data subjects' voices and lived experiences in mental healthcare provisioning. P81, a white woman with a mental health condition, expressed concerns about being mislabeled by emotion AI: "*Sometimes a system could tell you that you are at risk of something when you are really operating at a safe level. Each individual has a different pain tolerance level. Their own impressions should come first before being labeled.*" P81 both reflects concerns that emotion AI inferences generated without individual baselines would mislead treatment planning and that providers privileging their perceived objectivity may weaken data subjects' agency with automated systems that "*tell*" data subjects and providers about their mental health condition, rather than considering data subjects' voices that "*should come first before being labeled.*" Concerns that inferences would potentially invalidate or dismiss data subjects' voices is also an issue that some participants noted may arise if providers rely solely on emotion AI in practice, which we explore further in Section 4.2.2.

In sum, many participants shared concerns demonstrating how emotion AI may produce inaccurate inferences that negatively impact mental health assessments, diagnoses, and treatments, and highlight how data subjects' voices may be disregarded if providers place more value on emotion AI inferences than their own perspectives and lived experiences. Thus, participants' concerns elucidate the various ways emotion AI may be an unsuitable solution that may worsen present challenges of inaccuracy in mental healthcare.

*4.1.3 Perceived potential negative implications of emotion AI: increase providers' biases.* 14.2% of participants ($n$ = 56), 80.4% of whom identified with at least one marginalized identity, described how algorithmic biases laden in emotion AI could amplify mental healthcare providers' own biases, negatively affecting the mental healthcare data subjects receive. In contrast to the potential for emotion AI to mitigate providers' biases acknowledged in Section 4.1.1, participants were concerned that emotion AI could potentially intensify providers' biases when delivering mental healthcare.

Previous work has surfaced a range of algorithmic biases in emotion AI [54, 98, 123], a concern reflected by many participants who noted that emotion AI may encode biases that providers may then use to legitimate their own. P331, a white woman with multiple mental health conditions, described how emotion AI "*could be biased or based on stereotypes that could lead to incorrect information and harm by falsely associating traits with someone.*" P7, a transgender, non-binary white person with multiple mental health conditions, similarly mentioned concerns with potential biases within emotion AI: "*this system [emotion AI] could be built with gender and race biases that could harm myself and other individuals, especially if the system is rigid in what it deems 'unhealthy looking.'*" P331 and P7's concerns highlight the perceived potential for biased emotion AI inferences to perpetuate harmful stereotypes, and thus facilitate flawed mental healthcare provisions. P7 went on to describe the manifold ways biased emotion AI systems could affect the processes mental healthcare providers follow: "*Many times when women or femme presenting persons go into doctor's offices and don't look 'presentable enough' symptoms go overlooked or ignored but going in without makeup can also look 'sickly' and the system could detect that as illnesses that patient does not have. This is also something that could impact disabled (hard-of-hearing and deaf) people*

*and non-native speakers with the speech analysis portion since they would not be speaking with the typical speech patterns that this program is 'looking for.'*" P7 highlights how societal stereotypes reflected in emotion AI training data via biased inferences could potentially influence providers' decision-making regarding data subjects' mental health assessments, diagnoses, and treatments.

Some participants anticipated that providers may use emotion AI to defend their biases, which can harm mental healthcare provisions. P306, an Asian woman with a mental health condition, stated that "*we as humans are bad at understanding intersectionality so how do we expect to code a computer to understand it? I would hate for more discrimination to be a result of this,*" pointing to concerns that emotion AI would be incapable of understanding data subjects' multiple intersecting identities in a meaningful and non-reductive way when evaluating mental health. Thus, the technology's potential incapability to adequately account for data subjects' complex, intersecting identities may lead to "*more discrimination*" in healthcare, rather than combat it. The obscurity surrounding how emotion AI technologies are developed and the decisions developers make when building such systems complicates the expectations (as P306 notes) we may have about emotion AI. This lack of transparency and regulation further raises issues in the trade-off between accuracy and fairness in algorithmic systems [108], whereby the inferences and interventions made by emotion AI may lead to disparate negative consequences that are compounded for data subjects with intersecting marginalized identities.

P335, a white transgender man with multiple mental health conditions, mentioned similar concerns: "*There is also the given of human subjectivity still being there when the data is given to the healthcare provider, so it depends on the provider and their potential biases as well at the end of the day.*" Participants like P335 shared concern for the potential that emotion AI inferences could not only exploit providers' biases, but that providers' own biases could limit their ability to recognize potentially flawed algorithmic decisions. Thus, human-in-the-loop processes – often proposed to stem concerns with algorithmic decision-making [49] – would do little to address biased providers' limited ability to recognize potentially inaccurate results, which may be ultimately dangerous to data subjects' mental health and wellbeing in practice. Notably, participants' concerns regarding the potential for providers' biases to be exacerbated and its effects on mental healthcare provisioning with emotion AI use contrast emotion AI's proposed potential to mitigate biases in mental healthcare (described in Section 4.1.1).

### 4.2 Data subjects' voices

Many participants referred to the existing mental healthcare challenge whereby patients' voices are lost or ignored in provider interactions and care provisioning [26, 122, 156, 163, 183]. Participants acknowledged emotion AI could potentially facilitate conversations around mental health and amplify data subjects' voices [3]. Yet, participants also acknowledged how emotion AI could, instead, reduce or remove their voices in mental healthcare processes, hindering their ability to take part in decisions about their own mental health and wellbeing. In this section, we first unravel how participants' recognition of dismissed patient voices in traditional mental healthcare rendered emotion AI an enticing potential solution to amplify data subjects' voices. We then describe participants' concerns that emotion AI could in practice reduce or remove data subjects' voices and their ability to interact with mental healthcare providers altogether, harming data subjects' agency in mental healthcare processes. By foregrounding data subjects' perspectives, we show how

[3]Participants' responses regarding emotion AI use to amplify data subjects' voices align with the following factorial vignettes: using emotion AI in mental healthcare to develop an intelligent computer program to conduct mental health therapy; inferring moments patients may need emotional support and responding with an intelligent computer program.

implementing emotion AI to "solve" challenges related to patients' voices inadequately addresses this extant issue.

*4.2.1 Perceived potential impacts of emotion AI: facilitate data subjects' mental health information disclosures.* 9.6% of participants ($n$ = 38), 92.1% of whom identified with at least one marginalized identity, noted difficulties with openly communicating their mental health concerns with their providers, which emotion AI may help to resolve. P242, a Black woman who reported no mental health conditions, described how emotion AI systems "*could be beneficial if they actually provide emotional support, [as] it would feel less isolating and maybe like I was being seen if the program was acknowledging and backing up that my words and expressions actually indicate what I say they do and not what a medical professional (who is not listening anyway) has decided.*" P242's remarks highlight foundational issues within mental healthcare regarding how providers often neglect their patients' voices and gaslight their concerns [7] (a problem pervasive especially for marginalized communities and Black folks, in particular [7, 27, 167]), leaving patients feeling unheard, unseen, and isolated, and potentially resulting in detrimental mental healthcare provisions. Participants like P242 acknowledged emotion AI as a potential solution to address this problem by legitimating patient concerns that providers often ignore, *if* the systems in practice provided meaningful emotional support and supported data subjects' voices during mental healthcare processes. As a result, emotion AI was perceived to potentially promote mental health information disclosures between mental healthcare patients and their providers by "*backing up*" the information patients disclose. However, as the following Section 4.2.2 explores, emotion AI could, in effect, worsen the problem by dismissing, rather than lifting, patient voices.

*4.2.2 Perceived potential negative implications of emotion AI: reduce or remove data subjects' voices and interactions with providers.* Human voices and interactions are important aspects of healthcare in general [21, 55]. Some participants perceived how emotion AI use in this domain may threaten the inclusion of patients' voices in mental healthcare processes. Previously, in Section 4.2.1, participants noted that patients' voices may already be ignored in mental healthcare processes, which emotion AI could potentially mitigate. However, this section describes participants' perceptions that emotion AI could potentially reduce or completely diminish data subjects' voices *and* interactions with mental healthcare providers if providers gave more weight to emotion AI inferences over data subjects' own voices and lived experiences.

**Dismissing data subjects' voices due to prioritizing emotion AI inferences.** 4.6% of participants ($n$ = 18), 72.2% of whom identified with at least one marginalized identity, reported concerns regarding how mental healthcare providers may become heavily reliant on emotion AI over data subjects' own voices and lived experiences. P113, a white woman with a mental health condition, shared that emotion AI could negatively impact her "*if doctors place complete confidence in software and discount the information [she] may tell them if it doesn't support software.*" P113 points to the perceived potential for providers to place more value on emotion AI inferences over data subjects' own information regarding their mental health– privileging biomedical expertise [13, 146] that echoes histories of mental illness patients being discredited, discounted, and gaslit about their own mental health-related experiences in interactions with providers [27, 167]. To add, P87, a white man with a mental health condition, stated: "*There is a danger then if such systems become widespread, it will become very difficult to refute their diagnoses,*" highlighting a concern that data subjects would be unable to contest the inferences that providers may take at face value. These remarks point toward the perceived potential for emotion AI inferences to become an unyielding point of reference for decisions on data subjects' mental health, without adequate consideration for data subjects' lived experiences in decisions related to their mental health. Participants' concerns regarding a reduction in data subjects' voices relate to other concerns that may surface from

emotion AI use including implications for the accuracy of mental healthcare provisions (explicated in Section 4.1.2) and biases in mental healthcare processes (covered in Section 4.1.3) whereby data subjects lose the ability to dispute the provisions or processes facilitated by emotion AI use.

**Diminished interactions with providers in mental healthcare.** In addition to data subjects' voices potentially becoming far removed from mental healthcare processes, 12.9% of participants ($n$ = 51), 76.5% of whom identified with at least one marginalized identity, also noted the potential for emotion AI use to reduce or remove interactions between data subjects and providers. P8, a white man with a mental health condition, stated, "*I don't think it's a good idea for anyone except whoever's building and selling these systems to remove much more of the human from human medicine.*" In other words, the human element (e.g., patient-provider interaction) is a salient aspect of mental healthcare that, if removed, may only ultimately benefit those who create and profit from the technology rather than those in need of care and subjected to its use. P8's belief in the importance of the human element highlights the need for human involvement in mental healthcare processes (even as problems with provider-patient interactions may persist), as reducing or supplanting these processes with emotion AI may consequently lead to harmful mental healthcare provisions. For example, P152, a white woman who reported no mental health conditions, stated that "*relying too heavily on computer-assisted programs can lead to poor healthcare. It could be tempting to use these programs to allow providers to step too far back from the process.*" P152's remarks illustrate how over-reliance on emotion AI could potentially result in "*poor healthcare*" outcomes and diminished interactions with providers, which may have implications for inaccurate (described in Section 4.1.2) or biased (explicated in Section 4.1.3) mental healthcare assessments, diagnoses, and treatments.

Some participants described how specific emotion AI-enabled technology, such as a chatbot that may augment or replace providers' involvement in mental healthcare provisions, may produce or perpetuate harm to data subjects' mental health. P242 noted that "*chatting with a chatbot for mental health support, that is only capable of providing canned planned responses could cause me to feel isolated, invisible and lead to depression or self-harm.*" P242's response highlights the potential for emotion AI interactions that inappropriately respond to emotion inferences to fail to meet data subjects' mental health needs and expose them to psychological harm. To add, P50, a white woman with multiple mental health conditions, reported being "*concerned about the human element of mental health evaluations and treatments getting lost as human beings are social animals and we are better at reading one another than a computer can ever be. Psychological healing also takes place primarily within human relationships, not AI chatbots.*" P50's response further underscores the importance of human interaction to mental healthcare processes, as leaving high-stakes decisions and interpersonal connections up to algorithmic models and "*chatbots*" may be harmful and insensitive to data subjects' mental health needs. This view reflects previous work on how emotion recognition-enabled wellbeing interventions are perceived to provide inadequate care fundamentally because of the lack of human interaction involved [157].

Altogether, participants' anticipated impacts warn that emotion AI-enabled technologies that replace human patient-provider interactions (e.g., chatbots) may augment therapeutic interactions with adequate, non-human automated care to the extent that they become artificial, result in psychological harms (e.g., feelings of neglect and ill-treatment), and, more fundamentally, reduce or remove data subjects' voices and interaction with providers in mental healthcare processes.

### 4.3 Monitoring data subjects for potential harm

In this section, we highlight participants' perceptions about emotion AI use for monitoring potential harm toward oneself or others to improve existing harm-prevention efforts – a use case

commonly proposed in previous work [14, 30, 37, 50, 52, 91, 148, 160, 164, 179]. [4] While participants acknowledged the potential merits of this use, they also raised significant concerns surrounding how emotion AI may inaccurately predict self-harm or harm to others, dangerously impacting data subjects as a result.

*4.3.1 Perceived potential impacts of emotion AI: identify potential data subject self-harm or harm posed to others.* 1.5% of participants ($n$ = 6), a striking 83.3% of whom identified with at least one marginalized identity, acknowledged how emotion AI could be used to monitor and identify individuals who may harm themselves or others. P322, a white woman with multiple mental health conditions, mentioned "*I think the only way this could work is by possibly monitoring a dangerous person's social media, or offering links and hotlines when someone is in need of immediate support.*" P322's comments underline perceptions that emotion AI would only be useful in this case *if* it invasively monitored sites of rich personal data (e.g., social media behavior) which may offer a unique window into an individual's intimate thoughts and affairs – an emerging area in which emotion AI may be used (e.g., digital phenotyping [42, 83]). However, P322's response also demonstrates the perceived potential for emotion AI use to intrude on data subjects' personal lives outside of mental healthcare and to conflate data subjects' online activity with potentially harmful offline behavior that may be inaccurately interpreted as "*someone in need of immediate support.*"

P279, a white woman with a mental health condition, acknowledged this application may be beneficial for *some* individuals, noting that it may "*help severely mentally ill people who need monitoring to stay safe,*" but that "*otherwise, [emotion AI would be]* way too invasive." P279 highlights the perceived need to protect *other* individuals with mental health conditions from potential self-harm while raising privacy concerns that legitimate the over-surveillance of mentally ill people to keep them "safe." Relatedly, prior work on suicide risk prediction on Facebook emphasizes that real-world implementations of automated harm prevention requires monitoring *all* users to effectively identify potential harm [65, 71]. Thus, it is critical to consider how *all* data subjects' privacy may be compromised for the purpose of harm prediction and prevention, the willingness of data subjects to be subjected to such surveillance, and the consequences of potentially inaccurate and harmful interventions.

*4.3.2 Perceived potential negative implications of emotion AI: inaccurate identifications of self-harm or harm posed to others.* Though using emotion AI for harm prevention may alluringly promise data subjects' safety, its inferences may be inaccurate and have detrimental impacts on data subjects. This section draws from responses in Section 4.1.2 where 51.4% of participants ($n$ = 203), 76.8% of whom identified with at least one marginalized identity, highlighted concern regarding potentially inaccurate inferences. We note that our qualitative analysis broadly categorized inaccurate emotion AI inferences as a perceived potential result of emotion AI use. From this analysis, we found that some participants more specifically highlighted emotion AI's potential to falsely identify individuals at risk of harming themselves or others.

P135, a white man who reported no mental health conditions, shared, "*I really don't think they can [be beneficial]. I can be mad about something and the system may interpret that I will hurt someone when in reality, I just want to tell someone about what happened.*" P135 also shared concerns he may "*involuntarily be subjected to unnecessary help or even restraint if the system concluded [he] was at risk of hurting someone when in reality, [he] just wanted to vent and it could be over and done within a few minutes of blowing off steam.*" P135's remarks contest earlier sentiments

[4] Perceptions of emotion AI use to identify potential self-harm or harm posed to others relate to factorial vignettes that asked participants about using emotion AI to infer whether patients are at risk of harming themselves and inferring harm to others.

in Section 4.3.1 regarding emotion AI as a tool for monitoring and detecting potential harm. Instead, P135 highlights how through a lack of adequate contextual understanding, emotion AI may inaccurately identify data subjects in unsafe situations (i.e., conflating venting behavior with danger and risk) and consequently expose data subjects to harm from inappropriately excessive responses (e.g., forced "help" or restraint). P193, a white woman who reported no mental health conditions, shared similar concerns: "*If someone is potentially self-harming or wants to commit suicide, this program may not recognize that. Or it could falsely flag someone as such. Reporting this to health services could be detrimental to the patient.*" P193 raises the question of whether subjecting data subjects to invasive surveillance methods is warranted if it may fail to identify at-risk individuals on the one hand, and on the other hand, expresses concern that false inferences can also be detrimental to the identified individual. Reporting inaccurate harm predictions "*to health services,*" for instance, may unjustly lead to police intervention and involuntary commitment [142, 157] which can result in unjustified physical harm or brutality [147] that disproportionately impacts marginalized communities [69, 90]. Participants' concerns also relate to notions of consent (mentioned in Section 4.4.2) and dismissed data subject voices (referred to in Section 4.2.2) whereby data subjects may be forced under surveillance without the ability to contest inferences made by emotion AI and associated interventions.

*4.3.3 Perceived potential negative implications of emotion AI: negative wellbeing effects.* As a result of being monitored by emotion AI, 4.6% of participants ($n$ = 18), 83.3% of whom identified with at least one marginalized identity, anticipated being subjected to emotion AI would negatively impact their wellbeing. P360, a white transgender person who reported having a mental health condition, stated that the idea of emotion AI in mental healthcare "*makes [them] uncomfortable.*" P373, a Black man who reported no mental health conditions, also said, "*continuously using the systems may cause [him] anxiety,*" while P282, a white man who reported no mental health conditions, described more specifically that emotion AI "*could lower self-esteem, frighten, put on the defensive, or otherwise make matters worse for individuals.*" These responses demonstrate the potential for emotion AI-enabled patient monitoring to induce negative wellbeing effects (e.g., feelings of fear, hypervigilance, low self-esteem) arising from the surveillance of data subjects' intimate and personal emotions, contradicting emotion AI's purported use to improve wellbeing [123, 126, 134]. We cover related privacy violations and the harms that may surface from them (e.g., distressing emotional harm that is caused by privacy violations [34] more comprehensively in Section 4.4.3.

## 4.4 Involved parties' understanding and uses of mental health inferences

Various parties are involved in mental healthcare, namely mental healthcare providers who distribute care provisions; insurance companies who determine access to mental healthcare; academic researchers who may use mental healthcare data to advance knowledge about mental health; and patients who are most affected by these parties' involvement in their mental healthcare. This section analyzes how participants acknowledged emotion AI inferences to potentially enhance a broader understanding of mental health, specifically for the benefit of mental healthcare providers and academic researchers.[5] However, this acknowledged emotion AI use highlights the lack of adequate mental health understanding today, creating an appeal for emotion AI's promises to enhance such understanding. Yet, participants expressed their worries regarding the potential for involved parties,

[5]The perceived potential use of emotion AI for enhancing involved parties' understanding of mental health is in line with factorial vignettes that describe the purpose of giving healthcare providers increased understanding about patients through data-driven insights and to share emotion AI inferences with academic researchers to help them learn more about mental health, as part of a research partnership.

including providers and insurance companies, to misuse emotion AI inferences in ways that may create barriers to accessing (quality) mental healthcare and compromise data subjects' privacy.

*4.4.1 Perceived potential impacts of emotion AI: enhance involved parties' understanding of mental health.* Mental healthcare requires an understanding of patients' needs. Due to the complex range of mental health conditions and symptoms, it is difficult to fully understand and tend to patients' needs. 9.6% of participants ($n$ = 38), 78.9% of whom identified with at least one marginalized identity, acknowledged how emotion AI may be used to enhance involved parties' (i.e., providers and academic researchers) understanding of mental health, which may allow providers to better tend to data subjects' needs and for academic researchers to advance understanding of mental health. P253, a Black woman, stated that emotion AI "*would benefit [me] greatly as having more ways to assess mental/physical health would give healthcare providers a better understanding of the patients they deal with,*" illustrating a perceived need for resources that aid providers' understanding of data subjects' mental health to then provide sufficient care. Additionally, P285, a white man who reported no mental health conditions, said "*They [emotion AI] could provide another way of providing insight into what is going on with me, or, if being done for research the researcher could help practitioner better understand some aspect of their practice.*" P285 describes how emotion AI inferences could potentially be used in research to both advance new knowledge about mental health and generate insights into mental health that would enhance healthcare practitioners' understanding of and approach to improving mental healthcare. These perceived emotion AI uses for enhancing involved parties' understanding of mental health (which may or may not result in improved care) point to a deficient understanding of patients' mental health experiences that currently challenges the state of mental healthcare.

*4.4.2 Perceived potential negative implications of emotion AI: creating barriers to accessing quality) mental healthcare.* Although emotion AI use is acknowledged to potentially enhance involved parties' understanding of mental health broadly, involved parties' possession of data subjects' emotion AI inferences may lead to harmful data misuse that may worsen mental healthcare inaccessibility (reflecting prior work [81, 89]). 4.8% of participants ($n$ = 19), 57.9% of whom identified with at least one marginalized identity, described how involved parties' (i.e., providers, insurance companies) access to emotion AI inferences may hinder data subjects' mental healthcare quality and insurance coverage, negatively affecting data subjects' access to professional medical care. P368, a Black woman with a mental health condition, shared that "*[emotion AI] could cause healthcare providers to create biases about their clients and even drop them from their system altogether. Deeming them 'high risk' and refusing to cover them.*" P368 illuminates concerns that emotion AI inferences may influence providers' biases (described in Section 4.2.2), potentially leading to harmful decisions that limit data subjects' access to mental healthcare and jeopardize their wellbeing. Similarly, P50, a white woman with multiple mental health conditions, stated, "*As a neurodiverse person with mental health issues, I worry that the quality of care that I would receive from healthcare providers would decrease dramatically if this technology was put in place by health providers to cut costs.*" These concerns point to the potential for emotion AI to entrench mental healthcare inequalities, and may have been shaped by the larger context of healthcare algorithmic systems that determine risk differently between Black and white patients, with downstream disparate effects on insurance coverage and costs [9, 27, 132, 153]. Similarly, P117, a multi-racial woman with multiple mental health conditions, stated that she believed emotion AI "*could be misused to limit access to certain treatments or services*" while P51, a Latina with a mental health condition, stated, "*This seems like it could be misused by health insurance companies to decrease client support and increase costs for clients.*" Overall, participants shared concern that underlying profit motives would drive the premature adoption of potentially harmful emotion AI. Perhaps P159 said it best: *"Given the sorry state of AIs,*

*the baked-in biases, and the overcapitalization of healthcare, I can only see this being used to deny service as a means of controlling costs, increasing profits and sold to major ad networks as yet another profit center without our knowledge or consent.*"

These concerns demonstrate the perceived potential for mental healthcare systems and insurance companies to misuse emotion AI to restrict mental health provisions and increase mental healthcare costs, economically benefiting these involved parties at the expense of harming data subjects needing mental health services by making it difficult to afford or access quality care. Furthermore, they reflect perceived violations of contextual integrity [130] where data subjects' emotional information may be inappropriately and detrimentally removed from its intended use for mental healthcare provisions to facilitate information misuse, and underscore the importance of understanding the perspectives of people *with* mental illness(es) concerning how they may be adversely impacted by emotion AI in the high-stakes context of mental healthcare.

*4.4.3 Perceived potential negative implications of emotion AI: intruding on data subjects' privacy.* Many participants shared concerns about emotion AI violating their privacy. Participants wondered if and how emotion AI would be regulated and how the data emotion AI collects and infers would be handled to ensure data subjects' privacy is protected and secured in practice. Participants described privacy concerns associated with emotion AI data handling practices and data subjects' ability to meaningfully consent to the collection and sharing of their emotional information. Throughout this section, we map participants' perceptions of potential privacy intrusions to distinct privacy harms outlined by Citron and Solove [34].

51.4% of participants ($n$ = 203), 77.8% of whom identified with at least one marginalized identity, shared concerns about how mental healthcare providers using emotion AI could invade data subjects' privacy, leading to myriad privacy harms. P92, a white woman with a mental health condition, stated, "*Sometimes we don't want to reveal things about ourselves. This would make me feel very vulnerable and exposed,*" pointing to how emotion AI may constrain data subjects' agency in exercising whether and to what extent they disclose private and sensitive mental health and emotion-related information to their provider. P92's concerns also reflect potential autonomy and emotional harms associated with emotion AI use in mental healthcare [34] by challenging data subjects' freedom to make decisions regarding their own data, including exposing their vulnerable emotions and personal information to involved parties other than their healthcare provider(s). Respecting data subjects' autonomy to reveal their emotional information (or not) is particularly salient given its sensitivity and vulnerability to abuse [5, 159].

Some participants also asked various questions concerning if and how emotion AI in mental healthcare would be regulated. P230, a white woman with a mental health condition, asked: "*How will the data be held? Will it be deleted afterward? If sent in for research, how many others will witness my data?*" P230's questions reflect participants' considerable privacy concerns surrounding emotion AI in mental healthcare, including its potential to harm patient autonomy [34] as a result of opaque and unregulated emotion AI data handling practices.

Even if mental healthcare providers and emotion AI vendors were to implement privacy-preserving designs and strict security controls, it is important to note that participants remained concerned about the potential for data leakages. P149, a white man with a mental health condition, noted: "*Even where the healthcare provider is ensuring confidentiality, I think there should be discussion as to whether such readings could ever be turned over to, or subpoenaed, by law enforcement officials or courts, and if so, under what specific circumstances,*" expressing concern about the potential for courts and law enforcement to compel healthcare providers to share an individual's stored emotion AI data, which is particularly notable given the history of forced hospitalization of those with mental health conditions [142, 157]. In addition, P345, a non-binary white person with a mental

health condition, expressed concerns "*about the safety of this information [generated by emotion AI]*" and its potential to be commodified if it were "*used outside of the healthcare field such as by advertisers or companies.*" Similarly, P265 described: "*The [collection of data subjects'] images/videos are of concern because they may get into the wrong hands or could be used for facial recognition beyond the supposed purpose. Even if I sign a consent, I do not trust that this information will be used appropriately and once it gets out into the open you are at a loss.* As P265 highlights, data subjects' emotion AI-generated inferences could be re-purposed and may be leaked to third parties, exposing patients' sensitive emotional information "*into the open*" and outside their control. Despite existing safeguards that may guarantee higher standards of confidentiality and data protection if emotion AI is used in clinical settings (e.g., the Health Insurance Portability and Accountability Act of 1996[6]) P149 and P345's concerns highlight the perceived potential for external data sharing and reuse beyond the original purpose of adopting emotion AI, and for data leakages (e.g., subpoenas, data breaches) to expose patients' sensitive emotional information in invisible and uncontrollable ways.

Potential external emotion AI data sharing and reuse may also involve relationship harms [34] as the trust between data subjects and their providers may be negatively impacted due to compromised patient confidentiality associated with emotion AI use. Moreover, the privacy harms implicated by emotion AI use may extend to individuals beyond the intended data subject. As P149 described: "*There's also the bystander issue... how would such audio or video recording ensure that other people's privacy in my residence was protected?*" P149's concerns point to the potential impact that emotion AI may have on *others'* privacy (e.g., partners, children) that may not be directly monitored by healthcare providers using emotion AI but whose interpersonal privacy is nonetheless implicated, and may fall outside the scope of any existing safeguards (i.e., HIPAA) designed to protect individual *patient* health information.

In all, participants described their concerns about how involved parties within and beyond the mental healthcare context may obtain and misuse data subjects' emotion AI inferences in ways outside of their intended purpose and the control of data subjects, consequently exposing data subjects to a range of privacy harms.

## 5 DISCUSSION

The relational ethics lens we apply to the realm of emotion AI in mental health technologies in this study surfaced considerable concerns of emotion AI-enabled risk held by participants, from their perspectives as patients, that challenge leading discourses that credulously assume that emotion AI's integration into mental healthcare will benefit patients. By privileging data subjects' insights who could be subjected to and most affected by emotion AI use in mental healthcare, our findings warn that adopting emotion AI can introduce problems that worsen the very challenges it often promises to solve and do so in ways that could reinscribe the marginalization already experienced by non-white, minority gender, and mentally ill patients.

As we show, although participants acknowledged the merit of various ways emotion AI use may improve mental healthcare, they remain concerned that unintended consequences will introduce new or exacerbate challenges that effectively worsen aspects of mental healthcare: impair, instead of improve, already inadequate mental health provisions; reduce, instead of enhance, patient voices; harm, instead of protect, patient safety and wellbeing; and, in the name of promoting enhancements

[6]The Health Insurance Portability and Accountability Act of 1996 or HIPAA is a federal law meant to protect patients' sensitive health information from disclosures without patients' consent or knowledge. However, we note that HIPAA does not cover all digital health applications, even when used in coordination with patients' healthcare providers, and only covers certain entities (providers, insurance companies, and business associates). Thus, patients may unwittingly authorize the disclosure of their personal health information to third parties which may or may not be considered a covered entity, and thus, not covered by HIPAA [178].

to mental health understanding, facilitate privacy intrusions and create new barriers to quality mental healthcare access. Notably, these surfaced concerns were shared by a majority of participants with at least one marginalized identity along the dimensions of race/ethnicity, gender, and mental health status – a finding that underscores relational ethics' utility to center disproportionately impacted groups to identify and begin to combat algorithmic injustices [17]. While more research is needed to further investigate how distinct marginalities contribute to patient perceptions of emotion AI's impacts, our work nonetheless illustrates that individuals who lack the social power to make decisions concerning emotion AI's adoption in mental healthcare do not share the same enthusiasm for emotion AI as healthcare researchers, practitioners, and technologists.

While participants acknowledged potential emotion AI uses to address or mitigate present mental healthcare challenges, our results should not be misconstrued for positive attitudes toward emotion AI itself. Participants' perceived potential positive emotion AI uses were, in most cases, more closely tied to a shared agreement that said challenges in mental healthcare exist and need to be addressed than a shared perception that emotion AI would positively address them. Participants noted that emotion AI could be beneficial *if* it facilitated earlier and more accurate diagnoses (and consequently, faster access to resources and treatments); *if* it actually provided interventions patients found to enhance their agency and be emotionally supportive; *if* it effectively prevents harm to a degree that outweighs its compromise to patient privacy. Participants acknowledged potential benefits of emotion AI because they found the *outcome* it promised beneficial, not necessarily the *means* by which it aims to do so.

In contrast, participants shared deep concerns about emotion AI's potential to negatively impact mental healthcare. For example, non-white participants expressed concern that variations in emotional expression across cultures and intersecting identities would not be adequately considered by emotion AI, leading to inaccurate inferences and discriminatory consequences. Similarly, transgender participants shared concerns about gender biases in emotion AI. They worried that its use would have adverse consequences for patients who do not present themselves according to stereotypical gender norms. These concerns are grounded in existing research showing that automatic gender classifications often exclude minority genders [97], emotion recognition algorithms are less accurate for non-white groups [79, 153, 196], and more fundamentally, assumptions made in many algorithmic emotion classification schemes that assume emotions are universally expressed (despite evidence to the contrary) [10, 56, 173]. In practice, algorithmic choices that fail to adequately consider variations across dimensions of identity may exacerbate existing health disparities whereby the only beneficiaries of mental healthcare outcomes associated with accurate emotion inferences are those for whom algorithmic inferences are more accurate, at the expense of worsening that same outcome for minoritized groups. Regardless of machine accuracy, the use of emotion AI would generate troves of additional patient data that non-white, minoritized genders, and participants with mental illness particularly worried would enable cost-cutting measures and data sharing practices that would compromise patient privacy and safety and disproportionately impact the quality and accessibility of mental healthcare for minoritized groups. Even if privacy policies restrict the sharing of emotional inferences, any stored data could be leaked in data breaches – a particular concern for patients with mental illness and minoritized genders who may be exposed to harm if their identifiable information were leaked. Participants' numerous concerns illustrate how emotion AI's integration into mental healthcare in practice could unfairly distribute its potential benefits to privileged parties while adversely affecting patients in ways that restrict patients' freedom from bias, negatively impact patients' welfare, and hinder patients' autonomy – especially patients from minoritized communities.

Our use of speculative methods centers the perceptions of emotion AI's potential data subjects – including those who may be most at-risk of emotion AI-enabled harms – and privilege their lived

experiences as their expertise [106]. Thus, our study surfaces how emotion AI's integration into mental healthcare threatens to impinge, rather than promote, the "enduring human values" (i.e., human welfare, autonomy, freedom from bias) [66] of the individuals who would be subject to and directly impacted by its implementation. Moreover, our findings present an enormous challenge to research, development, and design efforts concerned with ensuring that the values of data subjects targeted by emotion AI are embedded into the technology: *What role, if any, can and should emotion AI play in realizing the future of mental healthcare that patients want, instead of perpetuating what is wrong with their current realities?*

In the following sections, we expand on our findings to guide a justice-oriented approach to this challenge, describing 1) how proposed emotion AI uses to mitigate existing challenges in mental healthcare are insufficient techno-solutions that may exacerbate the very problems emotion AI is implemented to address by introducing harms that may be obscured beneath illusions of progress toward improving mental healthcare; and 2) how said techno-solutions can further entrench injustices in mental healthcare by mapping participants' perceived negative impacts of emotion AI use to (in)justice frameworks and explicating the implications therein.

### 5.1 Emotion AI as a techno-solution and an illusion of progress toward improved mental healthcare

There are a plethora of mental healthcare challenges that patients *and* providers may face concerning seeking, receiving, and providing mental healthcare in the U.S., respectively. Regarding patients in particular, prior work describes some major challenges concerning patients' mental healthcare provisions [1, 33, 99, 113, 156, 174, 198], patients' voices and interactions with providers [26, 122, 163, 183], patient self-harm or harm posed to others [14, 37, 148, 160], and involved parties' understanding and uses of mental health information [80, 89]. Thus, emotion AI is a proposed *techno-solution* that mental healthcare systems, involved parties (e.g., providers and technologists) [100], and, as we show, patients may find appealing to address the aforementioned challenges through AI's perceived, yet exaggerated [10, 17, 158], qualities of objectivity and precision. Participants in our study acknowledged how emotion AI could be incorporated into mental healthcare processes to potentially "solve" issues of inaccurate, inefficient, and biased mental health assessments, diagnoses, and treatments; dismissed patient voices in mental healthcare procedures; patient safety; and involved parties' insufficient understanding of mental health.

Participants' perceptions that emotion AI use in mental healthcare may impact data subjects positions emotion AI as a techno-solution, or a technological intervention that "can unilaterally solve difficult social problems" [109]. *Techno-solutionism* relies on four central assumptions – 1) that the current state of affairs is deficient and that change is inherently positive; 2) that social phenomena and technological interventions can be seen as discrete; 3) that it is possible to cleanly delineate between undesirable and desirable social processes; and 4) that the advantages of technological interventions will be apparent to involved parties [67, 125]. Our findings reflect how emotion data subjects sometimes perceive AI as a techno-solution to the deficient state of mental healthcare, yet may be uncritically adopted by involved parties (e.g., technologists, mental healthcare providers, and insurance companies) in ways that may ignore harms posed to data subjects' wellbeing, privacy, and autonomy. Additionally, perceiving emotion AI as a techno-solution implies that underlying issues in mental healthcare are discrete and solvable via the implementation of technical solutions. By eliciting participants' perceptions of emotion AI use's impacts on data subjects in mental healthcare, we demonstrate how assumptions that emotion AI can "solve" existing challenges in mental healthcare (which are held by some participants in our study) may, ultimately, lead to neglecting the adverse effects of developing and implementing emotion AI on data subjects' personal and private emotional data, especially amongst those with marginalized identities. Although participants

perceived ways emotion AI may tackle existing mental healthcare challenges, they also noted how emotion AI use may in effect *exacerbate* these existing issues. Participants highlight that emotion AI use may increase inaccurate mental health provisions along with providers' biases; diminish data subjects' voices and interactions with providers in mental healthcare processes; inaccurately identify potential data subject self-harm or harm posed to others; and lead to involved parties' misuse of emotion AI inferences with consequences to (quality) mental healthcare access and data subjects' privacy. These perceived concerns of emotion AI use both 1) highlight the existing obstacles in mental healthcare including overworked [47, 127] or biased [113, 144] clinicians and the power dynamics dominating the patient-provider relationship [150] and overall healthcare domain; and 2) demonstrate the harms that may arise from attempting to solve such challenges with the techno-solution of emotion AI.

Participants' perspectives also highlight how emotion AI is not an inherently positive or clearly desirable technical solution for data subjects as mental healthcare patients. As we demonstrate, participants perceived emotion AI use in mental healthcare to facilitate and/or exacerbate the very problems that made emotion AI a potentially alluring techno-solution in the first place. For instance, some participants perceived how emotion AI-enabled chatbots can serve as a techno-solution that facilitates poor mental healthcare provisions, restricts patient autonomy, and diminishes patient-provider interactions which may lead to harmful mental healthcare outcomes and alienating data subjects, contesting prior work that advocates for the use of chatbots [112, 162, 184]. Thus, while participants acknowledged emotion AI as a techno-solution to various mental healthcare challenges (e.g., mitigating overworked, biased, and/or inattentive clinicians and recognize potential (self) harm), they also described how it may have detrimental repercussions on data subjects' mental healthcare at a larger scale such as by worsening mental healthcare provisions (concerning accuracy and bias), inauthentic care, and harmful opportunities for privacy intrusions and data misuse.

Perhaps most notably, participants called attention to how emotion AI as a techno-solution may further result in providers ignoring or disregarding data subjects' voices in mental healthcare processes, a mental healthcare problem echoed in prior work [26, 122, 156, 163, 183]. As such, diminishing or eliminating data subjects' voices, perspectives, and lived experiences from mental healthcare processes can be detrimental for both data subjects and the healthcare system by depending on emotion AI use. Diminishing data subjects' voices can also exacerbate the already documented problem of medical gaslighting, the bias that reveals itself in interpersonal interactions between patients and providers. Yet, scholars like Sebring [167] note that medical gaslighting is both interpersonal and resultant of embedded and historically unchallenged ideologies that undergird healthcare services. As such, medical gaslighting disproportionately impacts those who have been "othered" by medical establishments, namely women, transgender, queer, low-income, disabled people, and people of color [7, 27, 167]. Our findings show that data subjects may experience new or worsened medical gaslighting if providers apply emotion AI and prioritize its generated inferences to supplant their patient's claims regarding their own mental health conditions rather than involving emotion AI to support data subjects' own voices in these processes. Thus, medical gaslighting can also be a form of ontological or epistemic violence wherein biomedical expertise is privileged over patients' lived experiences [7].

Using emotion AI as a techno-solution that may adversely diminish data subjects' voices can also be troublesome for the healthcare system at large. Past scholarship notes that the inclusion of "patient voices" in healthcare provisions is crucial as they can help set research and treatment priorities, communicate the impact of disease, and report patient outcomes [51]. Additionally, including patients' voices is integral to the healthcare system's provisions of "patient-centered" treatments, which involves 1) patients as partners in the healthcare provision process, 2) providers' transparency about the processes they use for diagnosis and treatment, 3) providers considering

the diversity of patient populations, 4) providers considering the desired outcomes that are most relevant for patients, and 5) patients' ability to provide data about their own experiences with the healthcare system [143]. If emotion AI diminishes the "patient voice," it may preclude healthcare entities from engaging in patient-centered treatments overall.

However, even where healthcare systems involve data subjects through emotion AI use processes, our findings illustrate other various concerns that may arise from emotion AI use as a techno-solution to mental healthcare challenges. Participants highlighted the perceived potential for emotion AI to inaccurately identify self-harm or harm being posed to others. Inaccurate identifications may be harmful by risking one's livelihood and/or involuntarily subjecting data subjects to harmful interventions, such as police intervention or involuntary commitment [142, 157], which already disproportionately impact marginalized communities [69, 90]. To add, emotion AI use may lead to involved parties gaining access to and misusing data subjects' mental health data in ways that violate the contextual integrity of disclosures between patients and healthcare providers and consequently induce various privacy harms [34, 130]. Thus, we argue that emotion AI as a techno-solution to manifold mental healthcare challenges is an unfit approach that, even with technical or procedural fixes, may continue to adversely perpetuate the many issues it is set out to solve and that may be disproportionately felt by marginalized communities. Nonetheless, addressing concerns in contexts where emotion AI may be implemented requires multiple approaches to preventing and mitigating potential harm to data subjects that may still be deficient in addressing mental healthcare problems.

### 5.2 Injustices stemming from emotion AI use in mental healthcare

To further highlight participants' perceived impacts of emotion AI use in mental healthcare, we interpret our findings through the lens of justice frameworks in this section, showing how emotion AI use may facilitate distributive, procedural, and interactional injustices for data subjects. Justice frameworks can be used as lenses to understand and expose the deeper, underlying issues with emotion AI use and how techno-solutionist approaches are unfit to "solve" extant challenges in mental healthcare.[7]

Our findings provide valuable insights into how emotion AI use may impose various injustices that are crucial to further explore and consider during the emotion AI development and adoption. To add, our methodological choice to oversample for individuals with marginalized identities magnifies the voices of those who may experience injustices more greatly and otherwise be disregarded [123, 154, 157, 176, 185, 200]. Throughout our analysis, we noted that a majority of marginalized participants shared sentiments regarding both acknowledged potential emotion AI uses to mitigate existing mental healthcare challenges *and* to exacerbate them, demonstrating how both present mental healthcare problems and ways emotion AI use may exacerbate them may indeed be disproportionately felt by marginalized communities.

*5.2.1 Distributive injustice.* Cook & Hegtvedt [38] define *distributive justice* as fair allocation of "valued rewards, resources, rights, obligations, etc. to an array of recipients." Distributive (in)justice related to emotion AI use in mental healthcare represents the allocation of material (i.e., assessments, diagnoses, and treatments) and immaterial (i.e., health and wellbeing conditions) outcomes.

Our findings warn that emotion AI use could adversely affect the mental healthcare patients receive by facilitating even more inaccurate and biased material outcomes (e.g., mental healthcare assessments, diagnoses, and treatments) than they already face [1, 33, 99, 113, 156, 174, 198]. Moreover, our findings show how emotion AI might further promote unjust immaterial mental

[7]We maintain that the justice types reviewed above are practically linked and not necessarily mutually exclusive and, thus, it is important to consider the intersecting injustices that may arise from emotion AI use.

health outcomes by implementing surveillance methods that stoke anxiety and fear stemming from emotion AI's deep privacy intrusions and that enable third-party (e.g., healthcare systems, insurance companies, advertisers) misuses of patients' emotion inferences that expose patients to privacy harms [34] Thus, our findings indicate that emotion AI use cases to allocate mental healthcare provisions more efficiently and accurately may effectively exacerbate, rather than mitigate, distributive injustices for its potential data subjects.

*5.2.2 Procedural injustice.* Cook & Hegtedt define *procedural justice* as fair procedures where "despite what might be perceived as a fair or just distribution of outcomes, the procedures by which the distribution was arrived at may be defined as unjust or illegitimate" [38]. Procedural justice often involves "structural features of the decision-making process" [45]. Importantly, unjust *procedures* may shape the unjust *allocation* of resources – although decision-making processes may be just in a *distributive* sense, the *processes* underpinning these distributions may not be.

Participants described how emotion AI use may magnify providers' biases through emotion inferences that, reflecting existing demographic stereotypes in its training data, may be less accurate for marginalized groups. Healthcare providers' misguided confidence in the inference's accuracy and objectivity could then negatively influence providers' decisions regarding mental healthcare provisions for patients whose emotion inferences are less accurate. If emotion AI facilitates a fairer distribution of mental health outcomes for patients (i.e., more accurate and objective information about patient emotions) through processes involving biased algorithmic predictions that only advantage socially privileged identities, then this unjust procedure could lead to the unjust allocation of mental healthcare resources as well.

*5.2.3 Interactional injustice. Interactional justice* "refers to the quality of the interpersonal interaction between individuals" [45]. Scholars have also described the potential overlap between *procedural* and *interactional* justice but note that procedural justice typically refers to the formal aspects of a process. Interactional justice refers to more informal, social aspects [45]. Moreover, some scholarship parses interactional justice into separate but related constructs of interpersonal (i.e., dignity and respect for persons) and informational (in)justice (i.e., meaningful transparency of decision-making systems) [35, 117].

Participants shared concern for the potential that providers would take emotion inferences at face value and prioritize emotion inferences over data subjects' voices and lived experiences. By serving as a proxy for patients' emotional experiences to mediate interactions between providers and patients, the use of emotion inferences in mental healthcare may negatively impact mental healthcare provisions by supplanting data subjects' own feelings and perceptions and by excluding them from formal mental healthcare procedures. Thus, participants' concerns about the very use of emotion inferences to stand in for their emotional experiences reflects how emotion AI's integration into mental healthcare can promote interactional injustices that compromise dignity and respect for persons during mental healthcare provisions.

In sum, situating our findings through justice frameworks demonstrates that using emotion AI in mental healthcare is a potentially detrimental approach, or "techno-solution," to addressing data subjects' mental health needs and problems plaguing the mental healthcare system (e.g., inadequate mental healthcare provisions [1, 33, 99, 113, 156, 174, 198], dismissal of patient voices [21, 55]), which may be especially felt by marginalized communities [17, 123, 154, 157, 176, 185, 200]. As our findings suggest, implementing emotion AI technology in mental healthcare may adversely result in exacerbated extant mental healthcare problems. Applying justice frameworks to our findings further explicates how emotion AI could worsen, rather than alleviate, mental healthcare challenges in ways that facilitate injustices of mental healthcare patients. While the impulse to "re-design" or "de-bias" emotion AI through technical means may be an alluring approach, our findings related to

procedural and interactional injustices suggest that such approaches may, too, be insufficient as they may only address distributive justice concerns without effectively mitigating other concerns such as diminished patient voice and potential data misuse. Participants perceptions of potential emotion AI impacts in mental healthcare bring forth the salience of data subjects' perspectives to inform considerations (including those of involved parties such as providers and technologists) around emerging technologies such as emotion AI and to consider how technological implementations may induce tremendous harm to data subjects.

## 6 CONCLUSION

Emotion AI is increasingly infiltrating many aspects of our lives, including mental healthcare. This paper draws from a qualitative analysis of open-ended responses to survey questions ($n$ = 395) to investigate data subjects' perceptions of impacts that may emerge from emotion AI use in mental healthcare. Our analysis exposed various existing mental healthcare challenges that participants perceived emotion AI use may address *and* exacerbate. We found that participants perceived emotion AI use in mental healthcare to potentially impact data subjects by 1) addressing inadequate mental healthcare assessments, diagnoses, and treatments; 2) facilitating data subjects' mental health information disclosures; 3) identifying potential data subject self-harm or harm posed to others; and 4) enhancing involved parties' understanding of mental health. We complicate these perceived possibilities by highlighting participants' perceptions about ways that emotion AI use may negatively impact data subjects by exacerbating the same issues they were purported to solve: 1) increasing inaccurate assessments, diagnoses, and treatments along with providers' biases; 2) reducing or remove data subjects' voices and interactions with providers in mental healthcare processes; 3) inaccurately identifying potential data subject self-harm or harm posed to others with implications for negative wellbeing effects; 4) involved parties' misusing of emotion AI inferences with consequences to (quality) mental healthcare access and data subjects' privacy. Thus, participants' perspectives provide valuable insight into how emotion AI may not be a *desired* technical solution in mental healthcare, as perceived emotion AI uses are marked with a number of potential harms that may impact data subjects, especially those who hold marginalized identities.

This work contributes to a growing body of work regarding emotion AI's societal and ethical implications by examining data subjects' perspectives in the high-stakes context of mental healthcare. While prior work examines data subjects' attitudes in other contexts, attitudes are highly context-dependent and mental health is a pressing social and public health matter. We argue that emotion AI use is a techno-solution that provides an illusion of improved mental healthcare, and we show how its use is imbued with distributive, procedural, and interactional injustice, especially for marginalized data subjects. We urge future work to consider the harms that may arise from the development and deployment of emotion AI, which engaging with data subjects illustrates.

## ACKNOWLEDGMENTS

We acknowledge Dr. Karen Boyd for contributing to the study design and execution, including by analyzing patent applications (as part of a larger project with the last author) and providing us with a list of purposes claimed in emotion AI patent applications that informed the design of the survey vignettes. We thank Tillie Rosenberg for collaborating on data analysis and Serena Fan for contributing to codebook development. We are grateful to the National Science Foundation for sponsoring this work through award number 2020872 and CAREER award number 2236674. Last but not least, we thank the research participants for sharing their perceptions as well as the anonymous reviewers and associate chairs for their constructive feedback.

## REFERENCES


[1] Alexandra L. Adame, Matthew Morsey, Ronald Bassman, and Kristina Yates. 2017. A Brief History of the Psychiatric Survivor Movement. In *Exploring Identities of Psychiatric Survivor Therapists: Beyond Us and Them*, Alexandra L. Adame, Matthew Morsey, Ronald Bassman, and Kristina Yates (Eds.). Palgrave Macmillan UK, London, 33–53. https://doi.org/10.1057/978-1-137-58492-2_2

[2] Idris Adjerid, Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2013. Sleights of privacy: framing, disclosures, and the limits of transparency. In *Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUPS '13)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/2501604.2501613

[3] Muhammad Aurangzeb Ahmad, Arpit Patel, Carly Eckert, Vikas Kumar, and Ankur Teredesai. 2020. Fairness in Machine Learning for Healthcare. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*. ACM, Virtual Event CA USA, 3529–3530. https://doi.org/10.1145/3394486.3406461

[4] AMN Healthcare Center for Research. 2022. *2022 Survey of Physician Appointment Wait Times and Medicare and Medicaid Acceptance Rates*. Technical Report. AMN/Merritt Hawkins.

[5] Nazanin Andalibi and Justin Buss. 2020. The Human in Emotion Recognition on Social Media: Attitudes, Outcomes, Risks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376680

[6] Onur Asan, Alparslan Emrah Bayrak, and Avishek Choudhury. 2020. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *Journal of Medical Internet Research* 22, 6 (June 2020), e15154. https://doi.org/10.2196/15154

[7] Larry Au, Cristian Capotescu, Gil Eyal, and Gabrielle Finestone. 2022. Long covid and medical gaslighting: Dismissal, delayed diagnosis, and deferred treatment. *SSM - Qualitative Research in Health* 2 (Dec. 2022), 100167. https://doi.org/10.1016/j.ssmqr.2022.100167

[8] Andrew Ballard. 2019. Framing Bias in the Interpretation of Quality Improvement Data: Evidence From an Experiment. *International Journal of Health Policy and Management* 8, 5 (March 2019), 307–314. https://doi.org/10.15171/ijhpm.2019.08

[9] Kess L. Ballentine. 2019. Understanding Racial Differences in Diagnosing ODD Versus ADHD Using Critical Race Theory. *Families in Society* 100, 3 (July 2019), 282–292. https://doi.org/10.1177/1044389419842765 Publisher: SAGE Publications Inc.

[10] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. 2019. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest* 20, 1 (July 2019), 1–68. https://doi.org/10.1177/1529100619832930 Publisher: SAGE Publications Inc.

[11] Lisa Feldman Barrett, Batja Mesquita, and Maria Gendron. 2011. Context in Emotion Perception. *Current Directions in Psychological Science* 20, 5 (Oct. 2011), 286–290. https://doi.org/10.1177/0963721411422522

[12] Ivana Bartoletti. 2019. AI in Healthcare: Ethical and Privacy Challenges. In *Artificial Intelligence in Medicine*, David Riaño, Szymon Wilk, and Annette ten Teije (Eds.), Vol. 11526. Springer International Publishing, Cham, 7–10. https://doi.org/10.1007/978-3-030-21642-9_2 Series Title: Lecture Notes in Computer Science.

[13] Suze Berkhout and Juveria Zaheer. 2021. Digital Self-Monitoring, Bodied Realities: Re-Casting App-Based Technologies in First Episode Psychosis. *Catalyst: Feminism, Theory, Technoscience* 7, 1 (April 2021). https://doi.org/10.28968/cftt.v7i1.34101

[14] Alan L. Berman and Gregory Carter. 2020. Technological Advances and the Future of Suicide Prevention: Ethical, Legal, and Empirical Challenges. *Suicide and Life-Threatening Behavior* 50, 3 (June 2020), 643–651. https://doi.org/10.1111/sltb.12610

[15] Jaspreet Bhatia, Travis D. Breaux, Joel R. Reidenberg, and Thomas B. Norton. 2016. A Theory of Vagueness and Privacy Risk Perception. In *2016 IEEE 24th International Requirements Engineering Conference (RE)*. 26–35. https://doi.org/10.1109/RE.2016.20 ISSN: 2332-6441.

[16] Chloe E. Bird, Peter Conrad, Allen M. Fremont, and Stefan Timmermans. 2010. *Handbook of Medical Sociology, Sixth Edition*. Vanderbilt University Press. Google-Books-ID: DYCayq0Fp2AC.

[17] Abeba Birhane. 2021. Algorithmic injustice: a relational ethics approach. *Patterns* 2, 2 (Feb. 2021), 100205. https://doi.org/10.1016/j.patter.2021.100205

[18] Nadège Bourvis, Aveline Aouidad, Michel Spodenkiewicz, Giuseppe Palestra, Jonathan Aigrain, Axel Baptista, Jean-Jacques Benoliel, Mohamed Chetouani, and David Cohen. 2021. Adolescents with borderline personality disorder show a higher response to stress but a lack of self-perception: Evidence through affective computing. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 111 (Dec. 2021), 110095. https://doi.org/10.1016/j.pnpbp.2020.110095

[19] Karen Boyd and Nazanin Andalibi. 2022. Automated Emotion Recognition in the Workplace: How Proposed Technologies Reveal Potential Futures of Work.

[20] Virginia Braun, Victoria Clarke, Elicia Boulton, Louise Davey, and Charlotte McEvoy. 2021. The online survey as a *qualitative* research tool. *International Journal of Social Research Methodology* 24, 6 (Nov. 2021), 641–654. https://doi.org/10.1080/13645579.2020.1805550

[21] Judith Belle Brown, Moira Stewart, and Bridget L. Ryan. 2003. Outcomes of patient-provider interaction. In *Handbook of health communication*. Lawrence Erlbaum Associates Publishers, 141–161.

[22] Phil Brown, Crystal Adams, Rachel Morello-Frosch, Laura Senier, and Ruth Simpson. 2010. Health Social Movements: History, Current Work, and Future DIrections. In *Handbook of Medical Sociology, Sixth Edition*, Chloe E. Bird, Peter Conrad, Allen M. Fremont, and Stefan Timmermans (Eds.). Vol. 6th ed. Vanderbilt University Press, Nashville, 380–394. http://proxy.lib.umich.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=360701&site=ehost-live&scope=site

[23] Theodore M. Brown and Elizabeth Fee. 2014. Social Movements in Health. *Annual Review of Public Health* 35, 1 (March 2014), 385–398. https://doi.org/10.1146/annurev-publhealth-031912-114356

[24] Patrick Button, Eva Dils, Benjamin Harrell, Luca Fumarco, and David Schwegman. 2020. Gender Identity, Race, and Ethnicity Discrimination in Access to Mental Health Care: Preliminary Evidence from a Multi-Wave Audit Field Experiment. https://doi.org/10.3386/w28164

[25] Angela Cai and John Robst. 2016. The relationship between race/ethnicity and the perceived experience of mental health care. *American Journal of Orthopsychiatry* 86 (2016), 508–518. https://doi.org/10.1037/ort0000119 Place: US Publisher: Educational Publishing Foundation.

[26] Ángela Carbonell, José-Javier Navarro-Pérez, and Maria-Vicenta Mestre. 2020. Challenges and barriers in mental healthcare systems and their impact on the family: A systematic integrative review. *Health Social Care in the Community* 28, 5 (2020), 1366–1379. https://doi.org/10.1111/hsc.12968 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/hsc.12968.

[27] Chelsey R. Carter. 2022. Gaslighting: ALS, anti-Blackness, and medicine. *Feminist Anthropology* n/a, n/a (2022). https://doi.org/10.1002/fea2.12107 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/fea2.12107.

[28] Stacy M. Carter, Wendy Rogers, Khin Than Win, Helen Frazer, Bernadette Richards, and Nehmat Houssami. 2020. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *The Breast* 49 (Feb. 2020), 25–32. https://doi.org/10.1016/j.breast.2019.10.001

[29] Stevie Chancellor, Michael L. Birnbaum, Eric D. Caine, Vincent M. B. Silenzio, and Munmun De Choudhury. 2019. A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, Atlanta GA USA, 79–88. https://doi.org/10.1145/3287560.3287587

[30] Stevie Chancellor and Munmun De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *npj Digital Medicine* 3, 1 (March 2020), 43. https://doi.org/10.1038/s41746-020-0233-7

[31] Adam Mourad Chekroud, Ryan Joseph Zotti, Zarrar Shehzad, Ralitza Gueorguieva, Marcia K Johnson, Madhukar H Trivedi, Tyrone D Cannon, John Harrison Krystal, and Philip Robert Corlett. 2016. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry* 3, 3 (March 2016), 243–250. https://doi.org/10.1016/S2215-0366(15)00471-X

[32] Jin Chen, Cheng Chen, Joseph B. Walther, and S. Shyam Sundar. 2021. Do You Feel Special When an AI Doctor Remembers You? Individuation Effects of AI vs. Human Doctors on User Experience. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–7. https://doi.org/10.1145/3411763.3451735

[33] Tom A.C. Chrisp, Sharon Tabberer, Benjamin D. Thomas, and Wayne A. Goddard. 2012. Dementia early diagnosis: Triggers, supports and constraints affecting the decision to engage with the health care system. *Aging Mental Health* 16, 5 (July 2012), 559–565. https://doi.org/10.1080/13607863.2011.651794

[34] Danielle Keats Citron and Daniel J. Solove. 2022. Privacy Harms. *Boston University Law Review* 102, 3 (April 2022), 793–863.

[35] Jason A. Colquitt. 2001. On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology* 86 (2001), 386–400. https://doi.org/10.1037/0021-9010.86.3.386 Place: US Publisher: American Psychological Association.

[36] NYU Web Communications. 2022. Health Care is Increasingly Unaffordable for People with Employer-Sponsored Health Insurance—Especially Women. http://www.nyu.edu/content/nyu/en/about/news-publications/news/2022/december/JAMA-employer-sponsored-health-insurance

[37] Benjamin L. Cook, Ana M. Progovac, Pei Chen, Brian Mullin, Sherry Hou, and Enrique Baca-Garcia. 2016. Novel Use of Natural Language Processing (NLP) to Predict Suicidal Ideation and Psychiatric Symptoms in a Text-Based Mental Health Intervention in Madrid. *Computational and Mathematical Methods in Medicine* 2016 (2016), 1–8. https://doi.org/10.1155/2016/8708434

[38] Karen S. Cook and Karen A. Hegtvedt. 1983. Distributive Justice, Equity, and Equality. *Annual Review of Sociology* 9, 1 (Aug. 1983), 217–241. https://doi.org/10.1146/annurev.so.09.080183.001245

[39] Nicholas C. Coombs, Wyatt E. Meriwether, James Caringi, and Sophia R. Newcomer. 2021. Barriers to healthcare access among U.S. adults with mental health challenges: A population-based study. *SSM - Population Health* 15 (Sept. 2021), 100847. https://doi.org/10.1016/j.ssmph.2021.100847

[40] Patrick W. Corrigan, Benjamin G. Druss, and Deborah A. Perlick. 2014. The Impact of Mental Illness Stigma on Seeking and Participating in Mental Health Care. *Psychological Science in the Public Interest* 15, 2 (Oct. 2014), 37–70. https://doi.org/10.1177/1529100614531398

[41] Shanley Corvite, Kat Roemmich, Tillie Rosenberg, and Nazanin Andalibi. 2022. Data Subjects' Perspectives on Emotion Artificial Intelligence Use in the Workplace: A Relational Ethics Lens.

[42] Kaitlin L. Costello and Diana Floegel. 2020. "Predictive ads are not doctors": Mental health tracking and technology companies. *Proceedings of the Association for Information Science and Technology* 57, 1 (2020), e250. https://doi.org/10.1002/pra2.250 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pra2.250.

[43] Kate Crawford. 2021. Artificial Intelligence Is Misreading Human Emotion. https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/ Section: Technology.

[44] Carlos Crivelli and Alan J. Fridlund. 2018. Facial Displays Are Tools for Social Influence. *Trends in Cognitive Sciences* 22, 5 (May 2018), 388–399. https://doi.org/10.1016/j.tics.2018.02.006

[45] Russell Cropanzano, Cynthia A. Prehar, and Peter Y. Chen. 2002. Using Social Exchange Theory to Distinguish Procedural from Interactional Justice. *Group Organization Management* 27, 3 (Sept. 2002), 324–351. https://doi.org/10.1177/1059601102027003002

[46] Michele L Crossley and Nick Crossley. 2001. 'Patient' voices, social movements and the habitus; how psychiatric survivors 'speak out'. *Social Science Medicine* 52, 10 (May 2001), 1477–1489. https://doi.org/10.1016/S0277-9536(00)00257-4

[47] Daniel Novinson, Kaleb Erickson, and Abraham Kim. 2023. Clinicians Feel Increasingly Overworked, Even Amid a Waning Pandemic. https://opmed.doximity.com/articles/clinicians-feel-increasingly-overworked-even-amid-a-waning-pandemic :~:text=Overwork%20remains%20a%20persistent%20problem,into%20the%20COVID%2D19%20pandemic.

[48] Charles Darwin. 1872. *The expression of the emotions in man and animals.* (first ed.). London: John Murray.

[49] Maria De-Arteaga, Riccardo Fogliato, and Alexandra Chouldechova. 2020. A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. https://doi.org/10.1145/3313831.3376638

[50] Munmun De Choudhury and Scott Counts. 2013. Understanding affect in the workplace via social media. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 303–316. https://doi.org/10.1145/2441776.2441812

[51] Samera Dean, Jonathan M. Mathers, Melanie Calvert, Derek G. Kyte, Dolores Conroy, Annie Folkard, Sue Southworth, Philip I. Murray, and Alastair K. Denniston. 2017. "The patient is speaking": discovering the patient voice in ophthalmology. *British Journal of Ophthalmology* 101, 6 (June 2017), 700–708. https://doi.org/10.1136/bjophthalmol-2016-309955 Publisher: BMJ Publishing Group Ltd Section: Review.

[52] Mandar Deshpande and Vignesh Rao. 2017. Depression detection using emotion artificial intelligence. In *2017 International Conference on Intelligent Sustainable Systems (ICISS)*. 858–862. https://doi.org/10.1109/ISS1.2017.8389299

[53] Mari Devereaux, Kara Hartnett, Caroline Hudson, Alex Kacik, Gabriel Perna, Nona Tepper, and Brock E. W. Turner. 2023. What's ahead for digital health in 2023? https://digitalhealth.modernhealthcare.com/digital-health/digital-health-stakeholders-offer-their-2023-predictions Section: Digital Health.

[54] Artem Domnich and Gholamreza Anbarjafari. 2021. Responsible AI: Gender bias assessment in emotion recognition. (2021). https://doi.org/10.48550/ARXIV.2103.11436 Publisher: arXiv Version Number: 1.

[55] Christine W. Duclos, Mary Eichler, Leslie Taylor, Javan Quintela, Deborah S. Main, Wilson Pace, and Elizabeth W. Staton. 2005. Patient perspectives of patient–provider communication after adverse events. *International Journal for Quality in Health Care* 17, 6 (Dec. 2005), 479–486. https://doi.org/10.1093/intqhc/mzi065

[56] Paul Ekman. 1972. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology* 53, 4 (1972), 712. ISBN: 1939-1315 Publisher: American Psychological Association.

[57] Paul Ekman. 1984. *Expression and the nature of emotion.* Lawrence Erlbaum Associates, Hillsdale, NJ.

[58] Paul Ekman and Wallace V. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17, 2 (1971), 124–129. https://doi.org/10.1037/h0030377

[59] Steven Epstein. 1996. *Impure science: AIDS, activism, and the politics of knowledge.* Number 7 in Medicine and society. University of California press, Berkeley.

[60] Ettinger. 2022. Perspective | Teletherapy works, and it is vitally needed. *Washington Post* (Sept. 2022). https://www.washingtonpost.com/wellness/2022/05/23/teletherapy-mental-health-states-insurance/

[61] Amy L. Fairchild. 2006. Leprosy, Domesticity, and Patient Protest: The Social Context of a Patients' Rights Movement in Mid-Century America. *Journal of Social History* 39, 4 (2006), 1011–1043. http://www.jstor.org.proxy.lib.umich.edu/stable/3790239 Publisher: Oxford University Press.

[62] Paul Festor, Ibrahim Habli, Yan Jia, Anthony Gordon, A. Aldo Faisal, and Matthieu Komorowski. 2021. Levels of Autonomy and Safety Assurance for AI-Based Clinical Decision Systems. In *Computer Safety, Reliability, and Security. SAFECOMP 2021 Workshops (Lecture Notes in Computer Science)*, Ibrahim Habli, Mark Sujan, Simos Gerasimou, Erwin Schoitsch, and Friedemann Bitsch (Eds.). Springer International Publishing, Cham, 291–296. https://doi.org/10.1007/978-3-030-83906-2_24

[63] Casey Fiesler. 2021. Innovating Like an Optimist, Preparing Like a Pessimist: The Ethical Speculation and the Legal Imagination. *Colorado Technology Law Journal* 19, 1 (2021), 18.

[64] Janet Finch. 1987. The Vignette Technique in Survey Research. *Sociology* 21, 1 (Feb. 1987), 105–114. https://doi.org/10.1177/0038038587021001008 Publisher: SAGE Publications Ltd.

[65] Trehani M Fonseka, Venkat Bhat, and Sidney H Kennedy. 2019. The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors. *Australian New Zealand Journal of Psychiatry* 53, 10 (Oct. 2019), 954–964. https://doi.org/10.1177/0004867419864428

[66] Batya Friedman, Peter H. Kahn, Alan Borning, and Alina Huldtgren. 2013. Value Sensitive Design and Information Systems. In *Early engagement and new technologies: Opening up the laboratory*, Neelke Doorn, Daan Schuurbiers, Ibo van de Poel, and Michael E. Gorman (Eds.). Springer Netherlands, Dordrecht, 55–95. https://doi.org/10.1007/978-94-007-7844-3_4

[67] John Gardner and Narelle Warren. 2019. Learning from deep brain stimulation: the fallacy of techno-solutionism and the need for 'regimes of care'. *Medicine, Health Care and Philosophy* 22, 3 (Sept. 2019), 363–374. https://doi.org/10.1007/s11019-018-9858-6

[68] Vaibhav Garg, Kevin Benton, and L. Jean Camp. 2014. The Privacy Paradox: A Facebook Case Study. *SSRN Electronic Journal* (2014). https://doi.org/10.2139/ssrn.2411672

[69] Alicia Garza. 2016. *Who do you serve, who do you protect?: police violence and resistance in the United States*. Haymarket Books, Chicago, Illinois. OCLC: 952247161.

[70] Nina Gerber, Benjamin Reinheimer, and Melanie Volkamer. 2019. Investigating People's Privacy Risk Perception. *Proceedings on Privacy Enhancing Technologies* 2019, 3 (July 2019), 267–288. https://doi.org/10.2478/popets-2019-0047

[71] Norberto Nuno Gomes De Andrade, Dave Pawson, Dan Muriello, Lizzy Donahue, and Jennifer Guadagno. 2018. Ethics and Artificial Intelligence: Suicide Prevention on Facebook. *Philosophy Technology* 31, 4 (Dec. 2018), 669–684. https://doi.org/10.1007/s13347-018-0336-0

[72] Gregorio González-Alcaide, Mercedes Fernández-Ríos, Rosa Redolat, and Emilia Serra. 2021. Research on Emotion Recognition and Dementias: Foundations and Prospects. *Journal of Alzheimer's Disease* 82, 3 (Aug. 2021), 939–950. https://doi.org/10.3233/JAD-210096

[73] Gabriel Grill and Nazanin Andalibi. 2022. Attitudes and Folk Theories of Data Subjects on Transparency and Accuracy in Emotion Recognition. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–35. Publisher: ACM New York, NY, USA.

[74] Reshma Gupta, Leah Binder, and Christopher Moriates. 2020. Rebuilding Trust and Relationships in Medical Centers: A Focus on Health Care Affordability. *JAMA* 324, 23 (Dec. 2020), 2361. https://doi.org/10.1001/jama.2020.14933

[75] Oliver L Haimson, Nazanin Andalibi, Munmun De Choudhury, and Gillian R Hayes. 2018. Relationship breakup disclosures and media ideologies on Facebook. *New Media Society* 20, 5 (May 2018), 1931–1952. https://doi.org/10.1177/1461444817711402

[76] Oliver L. Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–35. https://doi.org/10.1145/3479610

[77] Marian Harbach, Sascha Fahl, and Matthew Smith. 2014. Who's Afraid of Which Bad Wolf? A Survey of IT Security Risk Awareness. In *2014 IEEE 27th Computer Security Foundations Symposium*. IEEE, Vienna, 97–110. https://doi.org/10.1109/CSF.2014.15

[78] Kate Hidalgo Bellows. 2022. Can Teletherapy Companies Ease the Campus Mental-Health Crisis? https://www.chronicle.com/article/can-teletherapy-companies-ease-the-campus-mental-health-crisis Section: News.

[79] Kasia Hitczenko, Henry R Cowan, Matthew Goldrick, and Vijay A Mittal. 2022. Racial and Ethnic Biases in Computational Approaches to Psychopathology. *Schizophrenia Bulletin* 48, 2 (March 2022), 285–288. https://doi.org/10.1093/schbul/sbab131

[80] Beatrix Hoffman. 2003. Health Care Reform and Social Movements in the United States. *American Journal of Public Health* 93, 1 (Jan. 2003), 75–85. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1447696/

[81] Sharona Hoffman and Andy Podgurski. 2013. The Use and Misuse of Biomedical Data: Is Bigger Really Better? *American Journal of Law Medicine* 39, 4 (Dec. 2013), 497–538. https://doi.org/10.1177/009885881303900401

[82] Natalie R. Holt, Debra A. Hope, Richard Mocarski, and Nathan Woodruff. 2023. The Often-Circuitous Path to Affirming Mental Health Care for Transgender and Gender-Diverse Adults. *Current Psychiatry Reports* 25, 3 (March 2023),

105–111. https://doi.org/10.1007/s11920-023-01410-2

[83] Kit Huckvale, Svetha Venkatesh, and Helen Christensen. 2019. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *npj Digital Medicine* 2, 1 (Sept. 2019), 1–11. https://doi.org/10.1038/s41746-019-0166-1 Number: 1 Publisher: Nature Publishing Group.

[84] Rhidian Hughes. 1998. Considering the Vignette Technique and its Application to a Study of Drug Injecting and HIV Risk and Safer Behaviour. *Sociology of Health Illness* 20, 3 (1998), 381–400. https://doi.org/10.1111/1467-9566.00107 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9566.00107.

[85] Nan D. Hunter. 2010. Rights Talk and Patient Subjectivity: The Role of Autonomy, Equality, and Participation Norms. *Wake Forest Law Review* 45, 5 (2010), 1525–1550. https://heinonline.org/HOL/P?h=hein.journals/wflr45&i=1535

[86] William James. 1948. What is emotion? 1884. In *Readings in the history of psychology.*, Wayne Dennis (Ed.). Appleton-Century-Crofts, East Norwalk, 290–303. https://doi.org/10.1037/11304-033

[87] Guillermina Jasso. 2006. Factorial Survey Methods for Studying Beliefs and Judgments. *Sociological Methods Research* 34, 3 (Feb. 2006), 334–423. https://doi.org/10.1177/0049124105283121 Publisher: SAGE Publications Inc.

[88] Bart Jones. 2022. Therapist shortage has parents struggling to get teens help as depression, suicides rise. https://www.newsday.com/news/health/therapy-teens-depression-suicide-wyotikol

[89] Timothy Judson, Mark Haas, and Tara Lagu. 2014. Medical Identity Theft: Prevention and Reconciliation Initiatives at Massachusetts General Hospital. *The Joint Commission Journal on Quality and Patient Safety* 40, 7 (July 2014), 291–AP1. https://doi.org/10.1016/S1553-7250(14)40038-2

[90] Kimberly Barsamian Kahn, Phillip Atiba Goff, J. Katherine Lee, and Diane Motamed. 2016. Protecting Whiteness: White Phenotypic Racial Stereotypicality Reduces Police Use of Force. *Social Psychological and Personality Science* 7, 5 (July 2016), 403–411. https://doi.org/10.1177/1948550616633505

[91] Haik Kalantarian, Khaled Jedoui, Peter Washington, Qandeel Tariq, Kaiti Dunlap, Jessey Schwartz, and Dennis P. Wall. 2019. Labeling images with facial emotion and the potential for pediatric healthcare. *Artificial Intelligence in Medicine* 98 (July 2019), 77–86. https://doi.org/10.1016/j.artmed.2019.06.004

[92] Jonathan W. Kanter, Daniel C. Rosen, Katherine E. Manbeck, Heather M. L. Branstetter, Adam M. Kuczynski, Mariah D. Corey, Daniel W. M. Maitland, and Monnica T. Williams. 2020. Addressing microaggressions in racially charged patient-provider interactions: a pilot randomized trial. *BMC Medical Education* 20, 1 (March 2020), 88. https://doi.org/10.1186/s12909-020-02004-9

[93] S Kapur, A G Phillips, and T R Insel. 2012. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry* 17, 12 (Dec. 2012), 1174–1179. https://doi.org/10.1038/mp.2012.105

[94] Harmanpreet Kaur, Daniel McDuff, Alex C. Williams, Jaime Teevan, and Shamsi T. Iqbal. 2022. "I Didn't Know I Looked Angry": Characterizing Observed Emotion and Reported Affect at Work. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–18. https://doi.org/10.1145/3491102.3517453

[95] Christopher J. Kelly, Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, and Dominic King. 2019. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine* 17, 1 (Oct. 2019), 195. https://doi.org/10.1186/s12916-019-1426-2

[96] Anne Kerr, Rosemary L. Hill, and Christopher Till. 2018. The limits of responsible innovation: Exploring care, vulnerability and precision medicine. *Technology in Society* 52 (Feb. 2018), 24–31. https://doi.org/10.1016/j.techsoc.2017.03.004

[97] Os Keyes. 2018. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 88:1–88:22. https://doi.org/10.1145/3274357

[98] Eugenia Kim, De'Aira Bryant, Deepak Srikanth, and Ayanna Howard. 2021. Age Bias in Emotion Detection: An Analysis of Facial Emotion Recognition Performance on Young, Middle-Aged, and Older Adults. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Virtual Event USA, 638–644. https://doi.org/10.1145/3461702.3462609

[99] Stuart A. Kirk and Herb Kutchins. 1988. Deliberate Misdiagnosis in Mental Health Practice. *Social Service Review* 62, 2 (June 1988), 225–237. https://doi.org/10.1086/644544

[100] Nikolaos Koutsouleris, Tobias U Hauser, Vasilisa Skvortsova, and Munmun De Choudhury. 2022. From promise to practice: towards the realisation of AI-informed mental health care. *The Lancet Digital Health* 4, 11 (Nov. 2022), e829–e840. https://doi.org/10.1016/S2589-7500(22)00153-4

[101] Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini, Arianna Manzini, Ilina Singh, and NeurOx Young People's Advisory Group. 2019. Can Your Phone Be Your Therapist? Young People's Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support. *Biomedical Informatics Insights* 11 (2019), 1178222619829083. https://doi.org/10.1177/1178222619829083

[102] Heather Kugelmass. 2016. "Sorry, I'm Not Accepting New Patients": An Audit Study of Access to Mental Health Care. *Journal of Health and Social Behavior* 57, 2 (June 2016), 168–183. https://doi.org/10.1177/0022146516647098

[103] Aaron Labbé. 2022. Council Post: Emotion AI: Why It's The Future Of Digital Health. https://www.forbes.com/sites/forbestechcouncil/2022/11/23/emotion-ai-why-its-the-future-of-digital-health/ Section: Innovation.
[104] Carl G. Lange. 1885. The Mechanism of the Emotions. *The Classical Psychologists* (1885), 672–684.
[105] Emily LaRosa and David Danks. 2018. Impacts on Trust of Healthcare AI. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New Orleans LA USA, 210–215. https://doi.org/10.1145/3278721.3278771
[106] Christopher A. Le Dantec, Erika Shehan Poole, and Susan P. Wyche. 2009. Values as lived experience: evolving value sensitive design in support of value discovery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1141–1150. https://doi.org/10.1145/1518701.1518875
[107] Min Kyung Lee and Katherine Rich. 2021. Who Is Included in Human Perceptions of AI?: Trust and Perceived Fairness around Healthcare AI and Cultural Mistrust. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. https://doi.org/10.1145/3411764.3445570
[108] Annie Liang, Jay Lu, and Xiaosheng Mu. 2022. Algorithmic Design: Fairness Versus Accuracy. In *Proceedings of the 23rd ACM Conference on Economics and Computation*. ACM, Boulder CO USA, 58–59. https://doi.org/10.1145/3490486.3538237
[109] Silvia Lindtner, Shaowen Bardzell, and Jeffrey Bardzell. 2016. Reconstituting the Utopian Vision of Making: HCI After Technosolutionism. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1390–1402. https://doi.org/10.1145/2858036.2858506
[110] Catherine Longworth. 2022. How advancing tech has shaped mental health in 2022. https://www.medicaldevice-network.com/features/how-advancing-tech-has-shaped-mental-health-in-2022/
[111] Ruth Ludwick, Marion Elizabeth Wright, Richard Allen Zeller, Dawn W. Dowding, William Lauder, and Janice Winchell. 2004. An Improved Methodology for Advancing Nursing Research: Factorial Surveys. *Advances in Nursing Science* 27, 3 (Sept. 2004), 224. https://journals.lww.com/advancesinnursingscience/Fulltext/2004/07000/An_Improved_Methodology_for_Advancing_Nursing.00007.aspx?casa_token=_AnESLB2tbMAAAAA:hg5wLigeL1U71WHASdKrz2dGbcWojJfWTnSx1no6B5horQJFDuSnZPWlSUlPoeOn5yKO5WFEU2GMalGBJ_CwKScR-i672g
[112] Kien Hoa Ly, Ann-Marie Ly, and Gerhard Andersson. 2017. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions* 10 (Dec. 2017), 39–46. https://doi.org/10.1016/j.invent.2017.10.002
[113] Ivy W. Maina, Tanisha D. Belton, Sara Ginzberg, Ajit Singh, and Tiffani J. Johnson. 2018. A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. *Social Science Medicine* 199 (Feb. 2018), 219–229. https://doi.org/10.1016/j.socscimed.2017.05.009
[114] Pauline Katharina Mantell, Annika Baumeister, Stephan Ruhrmann, Anna Janhsen, and Christiane Woopen. 2021. Attitudes towards Risk Prediction in a Help Seeking Population of Early Detection Centers for Mental Disorders—A Qualitative Approach. *International Journal of Environmental Research and Public Health* 18, 3 (Jan. 2021), 1036. https://doi.org/10.3390/ijerph18031036
[115] Kirsten Martin. 2016. Do Privacy Notices Matter? Comparing the Impact of Violating Formal Privacy Notices and Informal Privacy Norms on Consumer Trust Online. *The Journal of Legal Studies* 45, S2 (June 2016), S191–S215. https://doi.org/10.1086/688488 Publisher: The University of Chicago Press.
[116] Kirsten E. Martin. 2012. Diminished or Just Different? A Factorial Vignette Study of Privacy as a Social Contract. *Journal of Business Ethics* 111, 4 (Dec. 2012), 519–539. https://doi.org/10.1007/s10551-012-1215-8
[117] Suzanne S. Masterson, Zinta S. Byrne, and Hua Mao. 2005. Interpersonal and informational justice: Identifying the differential antecedents of interaction justice behaviors. In *What Motivates Fairness in Organizations?*
[118] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 72:1–72:23. https://doi.org/10.1145/3359174
[119] Daniel McDuff, Eunice Jun, Kael Rowan, and Mary Czerwinski. 2021. Longitudinal Observational Evidence of the Impact of Emotion Regulation Strategies on Affective Expression. *IEEE Transactions on Affective Computing* 12, 3 (July 2021), 636–647. https://doi.org/10.1109/TAFFC.2019.2961912 Conference Name: IEEE Transactions on Affective Computing.
[120] Andrew McStay. 2018. *Emotional AI: the rise of empathic media*. SAGE, Los Angeles.
[121] Andrew McStay. 2020. Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data Society* 7, 1 (Jan. 2020), 205395172090438. https://doi.org/10.1177/2053951720904386
[122] Francesca Mongelli, Penelope Georgakopoulos, and Michele T. Pato. 2020. Challenges and Opportunities to Meet the Mental Health Needs of Underserved and Disenfranchised Populations in the United States. *FOCUS* 18, 1 (Jan. 2020), 16–24. https://doi.org/10.1176/appi.focus.20190028 Publisher: American Psychiatric Publishing.

[123] Scott Monteith, Tasha Glenn, John Geddes, Peter C. Whybrow, and Michael Bauer. 2022. Commercial Use of Emotion Artificial Intelligence (AI): Implications for Psychiatry. *Current Psychiatry Reports* 24, 3 (March 2022), 203–211. https://doi.org/10.1007/s11920-022-01330-7

[124] Mordor Intelligence. 2022. Emotion Detection And Recognition (EDR) Market Share, Size | 2022 - 27 | Report, Trend. https://www.mordorintelligence.com/industry-reports/emotion-detection-and-recognition-edr-market

[125] Evgeny Morozov. 2013. *To Save Everything, Click Here: The Folly of Technological Solutionism.* PublicAffairs. Google-Books-ID: b3n8AgAAQBAJ.

[126] Salla Muuraiskangas, Marja Harjumaa, Kirsikka Kaipainen, and Miikka Ermes. 2016. Process and Effects Evaluation of a Digital Mental Health Intervention Targeted at Improving Occupational Well-Being: Lessons From an Intervention Study With Failed Adoption. *JMIR Mental Health* 3, 2 (May 2016), e13. https://doi.org/10.2196/mental.4465

[127] Ottar Mæstad, Gaute Torsvik, and Arild Aakvik. 2010. Overworked? On the relationship between workload and health worker performance. *Journal of Health Economics* 29, 5 (Sept. 2010), 686–698. https://doi.org/10.1016/j.jhealeco.2010.05.006

[128] Harris Meyer Kaiser Health News. 2022. Patients are turning to apps for therapy. But do digital mental health startups really help? https://www.latimes.com/business/story/2022-07-07/digital-mental-health-companies Section: Business.

[129] Guisi Ni, Weisong Shi, and Prashant Mahajan. 2014. Appurtenant: enhancing completeness and efficiency of bidirectional patient-physician communication using automatic speech recognition. In *Proceedings of the 2014 workshop on Mobile augmented reality and robotic technology-based systems - MARS '14.* ACM Press, Bretton Woods, New Hampshire, USA, 35–40. https://doi.org/10.1145/2609829.2609830

[130] Helen Nissenbaum. 2004. Privacy as Contextual Integrity. *Washington Law Review* 79 (2004), 119. https://heinonline.org/HOL/Page?handle=hein.journals/washlr79&id=129&div=&collection=

[131] Northwest ADA Center. 2018. Accessible Health Care. https://adata.org/factsheet/accessible-health-care

[132] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. (2019), 8.

[133] Songhee Oh, Jae Heon Kim, Sung-Woo Choi, Hee Jeong Lee, Jungrak Hong, and Soon Hyo Kwon. 2019. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. *Journal of Medical Internet Research* 21, 3 (March 2019), e12422. https://doi.org/10.2196/12422 Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

[134] Serena Oppenheim. 2019. How The Corporate Wellness Market Has Exploded: Meet The Latest Innovators In The Space. *Forbes* (June 2019). https://www.forbes.com/sites/serenaoppenheim/2019/06/11/how-the-corporate-wellness-market-has-exploded-meet-the-latest-innovators-in-the-space/?sh=5668dc065d91

[135] Trishan Panch, Tom J. Pollard, Heather Mattie, Emily Lindemer, Pearse A. Keane, and Leo Anthony Celi. 2020. "Yes, but will it work for my patients?" Driving clinically relevant research with benchmark datasets. *npj Digital Medicine* 3, 1 (June 2020), 1–4. https://doi.org/10.1038/s41746-020-0295-6 Number: 1 Publisher: Nature Publishing Group.

[136] Sun Young Park, Pei-Yi Kuo, Andrea Barbarin, Elizabeth Kaziunas, Astrid Chow, Karandeep Singh, Lauren Wilcox, and Walter S. Lasecki. 2019. Identifying Challenges and Opportunities in Human-AI Collaboration in Healthcare. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing.* ACM, Austin TX USA, 506–510. https://doi.org/10.1145/3311957.3359433

[137] Sarah Parvini. 2022. As L.A. County's mental health workers burn out, some weigh options. https://www.latimes.com/california/story/2022-12-20/la-me-mental-health-workers-los-angeles Section: California.

[138] Lindsey Patterson. 2012. Points of Access: Rehabilitation Centers, Summer Camps, and Student Life in the Making of Disability Activism, 1960-1973. *Journal of Social History* 46, 2 (Dec. 2012), 473–499. https://doi.org/10.1093/jsh/shs099

[139] Lindsey Patterson. 2018. The Disability Rights Movement in the United States. In *The Oxford Handbook of Disability History* (1 ed.), Michael Rembis, Catherine Kudlick, and Kim E. Nielsen (Eds.). Oxford University Press, 439–458. https://doi.org/10.1093/oxfordhb/9780190234959.013.0026

[140] Lindsey Marie Patterson. 2012. *The Right to Access: Citizenship and Disability, 1950-1973.* Ph. D. Dissertation. The Ohio State University. http://rave.ohiolink.edu/etdc/view?acc_num=osu1342310475

[141] Jessica K. Paulus and David M. Kent. 2020. Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. *npj Digital Medicine* 3, 1 (Dec. 2020), 99. https://doi.org/10.1038/s41746-020-0304-9

[142] Sachin R Pendse, Daniel Nkemelu, Nicola J Bidwell, Sushrut Jadhav, Soumitra Pathare, Munmun De Choudhury, and Neha Kumar. 2022. From Treatment to Healing:Envisioning a Decolonial Digital Mental Health. In *CHI Conference on Human Factors in Computing Systems (CHI '22).* Association for Computing Machinery, New York, NY, USA, 1–23. https://doi.org/10.1145/3491102.3501982

[143] Eleanor M. Perfetto, Elisabeth M. Oehrlein, Marc Boutin, Sarah Reid, and Eric Gascho. 2017. Value to Whom? The Patient Voice in the Value Discussion. *Value in Health* 20, 2 (Feb. 2017), 286–291. https://doi.org/10.1016/j.jval.2016.11.014

[144] Thomas V. Perneger and Thomas Agoritsas. 2011. Doctors and Patients' Susceptibility to Framing Bias: A Randomized Trial. *Journal of General Internal Medicine* 26, 12 (Dec. 2011), 1411–1417. https://doi.org/10.1007/s11606-011-1810-x

[145] Rosalind W. Picard. 1997. *Affective computing*. MIT Press, Cambridge, Mass.

[146] Martyn Pickersgill. 2019. Digitising psychiatry? Sociotechnical expectations, performative nominalism and biomedical virtue in (digital) psychiatric praxis. *Sociology of Health Illness* 41, S1 (2019), 16–30. https://doi.org/10.1111/1467-9566.12811 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9566.12811.

[147] Harold A. Pollack and Keith Humphreys. 2020. Reducing Violent Incidents between Police Officers and People with Psychiatric or Substance Use Disorders. *The ANNALS of the American Academy of Political and Social Science* 687, 1 (Jan. 2020), 166–184. https://doi.org/10.1177/0002716219897057

[148] Shraddha Pophale, Hetal Gandhi, and Anil Kumar Gupta. 2021. Emotion Recognition Using Chatbot System. In *Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications*, Vinit Kumar Gunjan and Jacek M. Zurada (Eds.). Vol. 1245. Springer Singapore, Singapore, 579–587. https://doi.org/10.1007/978-981-15-7234-0_54 Series Title: Advances in Intelligent Systems and Computing.

[149] Francisco A. Pujol, Higinio Mora, and Ana Martínez. 2019. Emotion Recognition to Improve e-Healthcare Systems in Smart Cities. In *Research Innovation Forum 2019*, Anna Visvizi and Miltiadis D. Lytras (Eds.). Springer International Publishing, Cham, 245–254.

[150] Robert W Putsch and Linda Pololi. 2004. Distributive Justice in American Healthcare: Institutions, Power, and the Equitable Care of Patients. *THE AMERICAN JOURNAL OF MANAGED CARE* 10 (2004), 9.

[151] Naila Rahman. 1996. Caregivers' Sensitivity to Conflict:. *Journal of Elder Abuse Neglect* 8, 1 (Aug. 1996), 35–47. https://doi.org/10.1300/J084v08n01_02 Publisher: Routledge _eprint: https://doi.org/10.1300/J084v08n01_02.

[152] Michael A. Rembis, Catherine Jean Kudlick, and Kim E. Nielsen. 2018. *The Oxford Handbook of Disability History*. Oxford University Press. Google-Books-ID: Sv5cDwAAQBAJ.

[153] Lauren Rhue. 2018. Racial Influence on Automated Perceptions of Emotions. https://doi.org/10.2139/ssrn.3281765

[154] Lauren Rhue. 2019. Affectively Mistaken? How Human Augmentation and Information Transparency Offset Algorithmic Failures in Emotion Recognition AI. https://doi.org/10.2139/ssrn.3492129

[155] Matt Richtel and Bee Trofort. 2022. 'Disruptive,' or Depressed? Psychiatrists Reach Out to Teens of Color. *The New York Times* (Dec. 2022). https://www.nytimes.com/2022/12/13/health/adolescents-mental-health-psychiatry.html

[156] Marc A. Rodwin. 1994. Patient Accountability and Quality of Care: Lessons From Medical Consumerism and the Patients' Rights, Women's Health and Disability Rights Movements. *American Journal of Law Medicine* 20, 1-2 (1994), 147–167. https://doi.org/10.1017/S009885880000647X Publisher: Cambridge University Press.

[157] Kat Roemmich and Nazanin Andalibi. 2021. Data Subjects' Conceptualizations of and Attitudes Toward Automatic Emotion Recognition-Enabled Wellbeing Interventions on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 308:1–308:34. https://doi.org/10.1145/3476049

[158] Kat Roemmich, Tillie Rosenberg, Serena Fan, and Nazanin Andalibi. 2023. Values in Emotion Artificial Intelligence Hiring Services: Technosolutions to Organizational Problems. (2023).

[159] Kat Roemmich, Florian Schaub, and Nazanin Andalibi. 2023. Emotion AI at Work: Implications for Workplace Surveillance, Emotional Labor, and Emotional Privacy. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–20. https://doi.org/10.1145/3544548.3580950

[160] John Rooksby, Alistair Morrison, and Dave Murray-Rust. 2019. Student Perspectives on Digital Phenotyping: The Acceptability of Using Smartphone Data to Assess Mental Health. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–14. https://doi.org/10.1145/3290605.3300655

[161] David C. Rubin and Jennifer M. Talarico. 2009. A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words. *Memory* 17, 8 (Nov. 2009), 802–808. https://doi.org/10.1080/09658210903130764

[162] Dinsa Sachan. 2018. Self-help robots drive blues away. *The Lancet Psychiatry* 5, 7 (July 2018), 547. https://doi.org/10.1016/S2215-0366(18)30230-X

[163] Nahal Salimi, Bryan Gere, William Talley, and Bridget Irioogbe. 2021. College Students Mental Health Challenges: Concerns and Considerations in the COVID-19 Pandemic. *Journal of College Student Psychotherapy* 0, 0 (Feb. 2021), 1–13. https://doi.org/10.1080/87568225.2021.1890298 Publisher: Routledge _eprint: https://doi.org/10.1080/87568225.2021.1890298.

[164] Samiha Samrose, Wenyi Chu, Carolina He, Yuebai Gao, Syeda Sarah Shahrin, Zhen Bai, and Mohammed Ehsan Hoque. 2019. Visual Cues for Disrespectful Conversation Analysis. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. 580–586. https://doi.org/10.1109/ACII.2019.8925440 ISSN: 2156-8111.

[165] Anvita Saxena, Ashish Khanna, and Deepak Gupta. 2020. Emotion Recognition and Detection Methods: A Comprehensive Survey. *Journal of Artificial Intelligence and Systems* 2, 1 (2020), 53–79. https://doi.org/10.33969/AIS.2020.21005

[166] Dagmar Schuller and Bjorn W. Schuller. 2018. The Age of Artificial Emotional Intelligence. *Computer* 51, 9 (Sept. 2018), 38–46. https://doi.org/10.1109/MC.2018.3620963

[167] Jennifer C. H. Sebring. 2021. Towards a sociological understanding of medical gaslighting in western health care. *Sociology of Health Illness* 43, 9 (2021), 1951–1964. https://doi.org/10.1111/1467-9566.13367 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9566.13367.

[168] Nassim Sheykholeslami. 2022. *Emotion AI in Mental Healthcare : How can affective computing enhance mental healthcare for young adults?* Student thesis. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-319800 Number Of Volumes: 2022:292 Publication Title: TRITA-EECS-EX Volume: Independent thesis Advanced level (degree of Master (Two Years)) 2022:292.

[169] Richard N. Shiffman, Bryant T. Karras, Sujai Nath, Laura Engles-Horton, and Geoffrey J. Corb. 1999. Pen-based, mobile decision support in healthcare. *ACM SIGBIO Newsletter* 19, 2 (Aug. 1999), 5–7. https://doi.org/10.1145/954507.954509

[170] Emma Silvers and Lesley McClurg. 2022. 'We're Drowning': Why Kaiser Mental Health Workers Are Striking. https://www.kqed.org/news/11923034/were-drowning-why-kaiser-mental-health-workers-are-striking

[171] Annie Snow, Julie Cerel, Diane N Loeffler, and Chris Flaherty. 2019. Barriers to Mental Health Care for Transgender and Gender-Nonconforming Adults: A Systematic Literature Review. *Health Social Work* 44, 3 (Aug. 2019), 149–155. https://doi.org/10.1093/hsw/hlz016

[172] Katta Spiel, Oliver L. Haimson, and Danielle Lottridge. 2019. How to do better with gender on surveys: a guide for HCI researchers. *Interactions* 26, 4 (June 2019), 62–65. https://doi.org/10.1145/3338283

[173] Luke Stark and Jesse Hoey. 2021. The Ethics of Emotion in Artificial Intelligence Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 782–793. https://doi.org/10.1145/3442188.3445939

[174] Derek H Suite, Robert La Bril, Annelle Primm, and Phyllis Harrison-Ross. 2007. Beyond Misdiagnosis, Misunderstanding and Mistrust: Relevance of the Historical Perspective in the Medical and Mental Health Treatment of People of Color. *JOURNAL OF THE NATIONAL MEDICAL ASSOCIATION* 99, 8 (2007).

[175] Žygintas Tamulis, Mindaugas Vasiljevas, Robertas Damaševičius, Rytis Maskeliunas, and Sanjay Misra. 2022. Affective Computing for eHealth Using Low-Cost Remote Internet of Things-Based EMG Platform. In *Intelligent Internet of Things for Healthcare and Industry*, Uttam Ghosh, Chinmay Chakraborty, Lalit Garg, and Gautam Srivastava (Eds.). Springer International Publishing, Cham, 67–81. https://doi.org/10.1007/978-3-030-81473-1_3

[176] The American Journal of Managed Care. 2006. Vulnerable Populations: Who Are They? *The American Journal of Managed Care (AJMC)* 12, 13 (Nov. 2006), 348–352. https://www.ajmc.com/view/nov06-2390ps348-s352 Publisher: MJH Life Sciences.

[177] The CONSORT-AI and SPIRIT-AI Steering Group. 2019. Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed. *Nature Medicine* 25, 10 (Oct. 2019), 1467–1468. https://doi.org/10.1038/s41591-019-0603-3

[178] Kim Theodos and Scott Sittig. 2021. Health Information Privacy Laws in the Digital Age: HIPAA Doesn't Apply. *Perspectives in Health Information Management* 18, Winter (2021), 1l.

[179] Laura K. Thompson, Margaret M. Sugg, and Jennifer R. Runkle. 2018. Adolescents in crisis: A geographic exploration of help-seeking behavior using data from Crisis Text Line. *Social Science Medicine* 215 (Oct. 2018), 69–79. https://doi.org/10.1016/j.socscimed.2018.08.025

[180] Graham Thornicroft, Diana Rose, and Aliya Kassam. 2007. Discrimination in health care against people with mental illness. *International Review of Psychiatry* 19, 2 (Jan. 2007), 113–122. https://doi.org/10.1080/09540260701278937 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09540260701278937.

[181] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science* 185, 4157 (Sept. 1974), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

[182] U.S. Department of Justice Civil Rights Division: Disability Rights Section. 2010. Americans with Disabilities Act: Access To Medical Care For Individuals With Mobility Disabilities. https://www.ada.gov/medcare_mobility_ta/medcare_ta.htm

[183] Ramyadarshni Vadivel, Sheikh Shoib, Sarah El Halabi, Samer El Hayek, Lamiaà Essam, Drita Gashi Bytyçi, Ruta Karaliuniene, Andre Luiz Schuh Teixeira, Sachin Nagendrappa, Rodrigo Ramalho, Ramdas Ransing, Victor Pereira-Sanchez, Chonnakarn Jatchavala, Frances Nkechi Adiukwu, and Ganesh Kudva Kundadak. 2021. Mental health in the post-COVID-19 era: challenges and the way forward. *General Psychiatry* 34, 1 (Feb. 2021), e100424. https://doi.org/10.1136/gpsych-2020-100424

[184] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S. Kashavan, and John Blake Torous. 2019. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *The*

*Canadian Journal of Psychiatry* 64, 7 (July 2019), 456–464. https://doi.org/10.1177/0706743719828977 Publisher: SAGE Publications Inc.

[185] Rosalie Waelen and Michał Wieczorek. 2022. The Struggle for AI's Recognition: Understanding the Normative Implications of Gender Bias in AI with Honneth's Theory of Recognition. *Philosophy Technology* 35, 2 (June 2022), 53. https://doi.org/10.1007/s13347-022-00548-w

[186] Ari Ezra Waldman. 2018. Privacy, Notice, and Design. *Stanford Technology Law Review* 21 (2018), 74. https://heinonline.org/HOL/Page?handle=hein.journals/stantlr21&id=74&div=&collection=

[187] Ari Ezra Waldman. 2020. Cognitive biases, dark patterns, and the 'privacy paradox'. *Current Opinion in Psychology* 31 (Feb. 2020), 105–109. https://doi.org/10.1016/j.copsyc.2019.08.025

[188] Lisa Wallander. 2009. 25 years of factorial surveys in sociology: A review. *Social Science Research* 38, 3 (Sept. 2009), 505–520. https://doi.org/10.1016/j.ssresearch.2009.03.004

[189] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan, and Feng Tian. 2021. "Brilliant AI Doctor" in Rural Clinics: Challenges in AI-Powered Clinical Decision Support System Deployment. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–18. https://doi.org/10.1145/3411764.3445432

[190] Qiaosi Wang, Shan Jing, David Joyner, Lauren Wilcox, Hong Li, Thomas Plötz, and Betsy Disalvo. 2020. Sensing Affect to Empower Students: Learner Perspectives on Affect-Sensitive Technology in Large Educational Contexts. In *Proceedings of the Seventh ACM Conference on Learning @ Scale*. ACM, Virtual Event USA, 63–76. https://doi.org/10.1145/3386527.3405917

[191] Meredith Whittaker, Meryl Alper, Olin College, Liz Kaziunas, and Meredith Ringel Morris. 2019. *Disability, Bias, and AI.* Technical Report. AINow.

[192] Diane Wiener, Rebecca Ribeiro, and Kurt Warner. 2009. Mentalism, disability rights and modern eugenics in a 'brave new world'. *Disability Society* 24, 5 (Aug. 2009), 599–610. https://doi.org/10.1080/09687590903010974 Publisher: Routledge _eprint: https://doi.org/10.1080/09687590903010974.

[193] Jenna Wiens, Suchi Saria, Mark Sendak, Marzyeh Ghassemi, Vincent X. Liu, Finale Doshi-Velez, Kenneth Jung, Katherine Heller, David Kale, Mohammed Saeed, Pilar N. Ossorio, Sonoo Thadaney-Israni, and Anna Goldenberg. 2019. Do no harm: a roadmap for responsible machine learning for health care. *Nature Medicine* 25, 9 (Sept. 2019), 1337–1340. https://doi.org/10.1038/s41591-019-0548-6

[194] Cathleen E. Willging, Melina Salvador, and Miria Kano. 2006. Brief Reports: Unequal Treatment: Mental Health Care for Sexual and Gender Minority Groups in a Rural State. *Psychiatric Services* 57, 6 (June 2006), 867–870. https://doi.org/10.1176/ps.2006.57.6.867 Publisher: American Psychiatric Publishing.

[195] Yunyu Xiao, Julie Cerel, and J. John Mann. 2021. Temporal Trends in Suicidal Ideation and Attempts Among US Adolescents by Sex and Race/Ethnicity, 1991-2019. *JAMA Network Open* 4, 6 (June 2021), e2113513. https://doi.org/10.1001/jamanetworkopen.2021.13513

[196] Tian Xu, Jennifer White, Sinan Kalkan, and Hatice Gunes. 2020. Investigating Bias and Fairness in Facial Expression Recognition. In *Computer Vision – ECCV 2020 Workshops*, Adrien Bartoli and Andrea Fusiello (Eds.). Vol. 12540. Springer International Publishing, Cham, 506–523. https://doi.org/10.1007/978-3-030-65414-6_35 Series Title: Lecture Notes in Computer Science.

[197] Eric Zeng, Shrirang Mare, and Franziska Roesner. 2017. End User Security & Privacy Concerns with Smart Homes. In *Proceedings of the Thirteenth Symposium on Usable Privacy and Security*. Santa Clara, CA, USA.

[198] Colin A. Zestcott, Irene V. Blair, and Jeff Stone. 2016. Examining the presence, consequences, and reduction of implicit bias in health care: A narrative review. *Group Processes Intergroup Relations* 19, 4 (July 2016), 528–542. https://doi.org/10.1177/1368430216642029

[199] John Zimmerman and Jodi Forlizzi. 2017. Speed Dating: Providing a Menu of Possible Futures. *She Ji: The Journal of Design, Economics, and Innovation* 3, 1 (March 2017), 30–50. https://doi.org/10.1016/j.sheji.2017.08.003

[200] Annuska Zolyomi and Jaime Snyder. 2021. Social-Emotional-Sensory Design Map for Affective Computing Informed by Neurodivergent Experiences. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 77:1–77:37. https://doi.org/10.1145/3449151

# A APPENDICES

## A FACTORIAL VIGNETTES FOR THE HEALTHCARE CONTEXT

The 14 purposes for which emotion AI may be deployed and informed our survey design are **bolded**. The 14 purposes were repeated twice to vary by two ways providers may automatically detect patients' emotional state, resulting in 28 factorial vignettes which are included below.

(1) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring the mental health state of patients. Inferences of an individual's mental health will not be made; only at a group level**.
(2) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) from a microphone, such as your heart rate to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring the mental health state of patients individually**.
(3) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of diagnosing mental illness in patients earlier than otherwise possible**.
(4) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of diagnosing neurological disorders, such as dementia or ADHD, in patients earlier than otherwise possible**
(5) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring patients in need of wellbeing support**.
(6) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of developing an intelligent computer program, such as a chat bot, that can conduct mental health therapy with patients, including you. Your information would be used to help test and train this program**.
(7) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring moments patients may be in need of emotional support, and responding with an intelligent computer program designed to help patients improve their wellbeing, such as offering wellbeing tips.**
(8) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of sharing that information with academic researchers to help them learn more about mental health, as part of a research partnership**.

(9) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of giving healthcare provider s) increased understanding about patients through data-driven insights**.

(10) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of automatically alerting your healthcare provider s) when patients may need support, including you**.

(11) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring whether patients are at risk of harming themselves**.

(12) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring whether patients are at risk of harming others**.

(13) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of avoiding human judgment and subjectivity present in ways patients typically provide this information, such as a self-report or through observation by your healthcare provider s)**.

(14) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses records of what you say (either verbally or written/typed) and how you say it (such as your speed or tone when saying it) to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of assessing the overall health of individual patients**.

(15) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring the mental health state of patients. Inferences of an individual's mental health will not be made; only at a group level**.

(16) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions,, such as your heart rate to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring the mental health state of patients individually**.

(17) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities

and device use, **for the purpose of diagnosing mental illness in patients earlier than otherwise possible**.

(18) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of diagnosing neurological disorders, such as dementia or ADHD, in patients earlier than otherwise possible**

(19) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring patients in need of wellbeing support**.

(20) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of developing an intelligent computer program, such as a chat bot, that can conduct mental health therapy with patients, including you. Your information would be used to help test and train this program**.

(21) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring moments patients may be in need of emotional support, and responding with an intelligent computer program designed to help patients improve their wellbeing, such as offering wellbeing tips**.

(22) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of sharing that information with academic researchers to help them learn more about mental health, as part of a research partnership**.

(23) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of giving healthcare provider s) increased understanding about patients through data-driven insights**.

(24) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of automatically alerting your healthcare provider s) when patients may need support, including you**.

(25) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring whether patients are at risk of harming themselves**.

(26) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of inferring whether patients are at risk of harming others**.

(27) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of avoiding human judgment and subjectivity present in ways patients typically provide this information, such as a self-report or through observation by your healthcare provider s)**.

(28) As a patient, rate your willingness to be the target of a software used by your healthcare provider(s) that uses images or video of what you look like, based on your facial expressions, to automatically infer your emotional state(s) based on information from your daily activities and device use, **for the purpose of assessing the overall health of individual patients**.

## B OPEN-ENDED QUESTIONS

(1) In what ways, if any, do you think these systems could benefit you? Please describe and provide examples and as much detail as you are comfortable with.
(2) In what ways, if any, do you think these systems could harm you or have other undesired impacts on you? Please describe and provide examples and as much detail as you are comfortable with.
(3) What other concerns, if any, do you have about these systems? Please describe and provide examples and as much detail as you are comfortable with.
(4) In what ways, if at all, do aspects of who you are (for example, your race/ethnicity, gender, sexuality, employment status, class, education, mental health conditions, physical health conditions, or any other features of your identity) shape your responses to the use of computer programs to infer your emotional states?

## C BREAKDOWN OF THE SAMPLE INCLUDED IN THIS PAPER

| Sample | Number of participants, |
|---|---|
| **Representative sample** | 289 |
| **Mental health oversample*** | 37 |
| **Gender oversample**** | |
| Trans | 6 |
| Non-binary | 26 |
| Trans, non-binary | 2 |
| **Race/ethnicity oversample***** | |
| African-American or Black | 11 |
| Asian-American | 1 |
| East Asian | 2 |
| Hispanic or Latino/a/x | 11 |
| Indigenous American or First Nations | 1 |
| Multi-racial | 9 |
| Total participants | 395 |

Table 3. Full breakdown of the sample included in this paper.

*Participants were asked "Please describe your mental health status. Select all the apply." to the following options: *I have a mental health condition and it has not been formally diagnosed; I have a mental health condition that has been formally diagnosed; I am being treated for a mental health condition, and that treatment includes medication; I am being treated for a mental health condition, not with medication; I do not have a mental health condition; I used to have a mental health condition and I no longer do; I have multiple mental health conditions. Some are diagnosed, some are not; I have multiple mental health conditions. I take medication for some, and do not for others.*

**Participants were asked "Please describe your gender. Select all that apply." to the following options: *Woman; Man; Trans; Non-binary; Prefer not to disclose; Prefer to self-describe (open-ended textbox).* These options were selected according to [172].

***Participants were asked "Please describe your race/ethnicity. Select all that apply." to the following options: *African; African-American or Black; Asian-American; East Asian; Hispanic or Latino/a/x; Indigenous American or First Nations; Middle Eastern; South Asian; Southeast Asian; White; Not listed; please specify (open-ended textbox); Prefer not to answer.*

## D BREAKDOWN OF PARTICIPANTS' DEMOGRAPHICS

*Participant age ranges were provided by the recruitment service Prolific and validated with a prescreening survey questions that asked participants' year of birth.

**Participants were asked "Please indicate your current employment status. Select all that apply." to the following options: *Employed Full-Time; Employed Part-Time; Looking for work; Not in the paid workforce (retired, full-time caregiving, full-time student, etc); Other (open-ended textbox)*

***Participants were asked "What is the highest level of school you have completed or the highest degree you have received?" to the following options: *No formal school; Some grade school; High school graduate (high school diploma or equivalent including GED); Some college; Technical, vocational, or trade school; Associate degree in college (2-year); Bachelor's degree in college (4-year); Master's degree; Professional degree (JD, MD); Doctoral degree*

****Participants were asked "Please describe your mental health status. Select all that apply" to the following options: *I have a mental health condition and it has not been formally diagnosed; I have a mental health condition that has been formally diagnosed; I am being treated for a mental health condition, and that treatment includes medication; I am being treated for a mental health condition, not with medication; I do not have a mental health condition; I used to have a mental health condition and I no longer do; I have multiple mental health conditions. Some are diagnosed, some are not; I have multiple mental health conditions. I take medication for some, and do not for others*

Received January 2023; revised July 2023; accepted November 2023

| Demographics | Number of participants, |
|---|---|
| **Gender** | |
| Woman | 202 |
| Man | 364 |
| Non-binary | 34 |
| Trans | 11 |
| **Race/ethnicity** | |
| African | 5 |
| African-American or Black | 62 |
| Asian-American | 27 |
| East Asian | 27 |
| Hispanic or Latino/a/x | 39 |
| Indigenous American or First Nations | 8 |
| Middle Eastern | 3 |
| Not listed | 7 |
| South Asian | 2 |
| Southeast Asian | 5 |
| White | 269 |
| **Age*** | |
| 18-24 | 96 |
| 25-34 | 83 |
| 35-44 | 61 |
| 45-54 | 50 |
| 55-64 | 58 |
| 65+ | 45 |
| **Employment status**** | |
| Employed full-time | 176 |
| Employed part-time | 61 |
| Not in the paid workforce | 100 |
| Looking for work | 47 |
| Other | 25 |
| **Highest level of education or degree completed***** | |
| Some grade school | 2 |
| High school graduate | 52 |
| Some college | 97 |
| Technical, vocational, or trade school | 5 |
| Associate degree in college | 41 |
| Bachelor's degree in college | 115 |
| Master's degree | 64 |
| Professional degree | 14 |
| Doctoral degree | 5 |
| **Mental health status****** | |
| Current or past lived experience with mental illness | 198 |
| Does not have current or past lived experience with mental illness | 197 |

Table 4. Note: Some percentages may add up to more than our sample number of 395 because participants could be in multiple gender and race/ethnicity categories and experiencing more than one employment event at once. Additionally 3 participants did not report their age.