Attitudes and Folk Theories of Data Subjects on Transparency and Accuracy in Emotion Recognition

GABRIEL GRILL, University of Michigan, USA
NAZANIN ANDALIBI, University of Michigan, USA

The growth of technologies promising to infer emotions raises political and ethical concerns, including concerns regarding their accuracy and transparency. A marginalized perspective in these conversations is that of data subjects potentially affected by emotion recognition. Taking social media as one emotion recognition deployment context, we conducted interviews with data subjects (i.e., social media users) to investigate their notions about accuracy and transparency in emotion recognition and interrogate stated attitudes towards these notions and related folk theories. We find that data subjects see accurate inferences as uncomfortable and as threatening their agency, pointing to privacy and ambiguity as desired design principles for social media platforms. While some participants argued that contemporary emotion recognition must be accurate, others raised concerns about possibilities for contesting the technology and called for better transparency. Furthermore, some challenged the technology altogether, highlighting that emotions are complex, relational, performative, and situated. In interpreting our findings, we identify new folk theories about accuracy and meaningful transparency in emotion recognition. Overall, our analysis shows an unsatisfactory status quo for data subjects that is shaped by power imbalances and a lack of reflexivity and democratic deliberation within platform governance.

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

Additional Key Words and Phrases: algorithm; folk theories; social media; emotion recognition; affect recognition; emotional artificial intelligence; transparency; accuracy; ethics

ACM Reference format:

Gabriel Grill, and Nazanin Andalibi. 2022. Attitudes and Folk Theories of Data Subjects on Transparency and Accuracy in Emotion Recognition. In *Proceedings of the ACM on Human-Computer Interaction,* Vol. 6, CSCW1, Article 78 (April 2022), 35 pages, https://doi.org/10.1145/3512925

1 INTRODUCTION

Emotion recognition and emotional artificial intelligence (AI) technologies have become widely used, and adoption is expected only to grow [6]. However, algorithmic inferences of emotions and affect are controversial. Several uses have been strongly critiqued, such as the Facebook

Author's addresses: Gabriel Grill, ggrill@umich.edu, University of Michigan, University of Michigan, 105 S State St, Ann Arbor, MI, 48109; Nazanin Andalibi, andalibi@umich.edu, University of Michigan, University of Michigan, 105 S State St, Ann Arbor, MI, 48109.

2573-0142/2022/04-78 $15.00

https://doi.org/10.1145/3512925

"emotional contagion" study and psychographic profiling by Cambridge Analytica [129]. Previous work has highlighted how emotions are perceived as private and personal in interactions with technology [6]. Simultaneously, emotions are understood as social, communicative, unstable, and cultural [129]. Overall, whether algorithms can truly capture emotions and the ethical permissibility of emotion recognition are contested [42,130]. Despite these concerns, technologies to detect and predict people's emotions based on online data are widely deployed in various domains. For example, companies use emotion recognition on social media to assess the success of advertising campaigns [104], academics employ it to generate scientific knowledge [131], political parties use it to understand public opinion and emotions for elections [130,139], and governments employ it for security purposes [130]. Missing from discourse regarding emotion recognition's societal impact are the perspectives of the people most impacted by it, and how they understand its key qualities (e.g., accuracy, transparency) [130]. This study is concerned with social media users' conceptions of emotion recognition technologies employed on social media platforms that produce and commercialize knowledge about users.

We understand emotion recognition to be an algorithmic assemblage [32,68.94,125], i.e., a set of components and processes implicated in broader algorithmic systems such as data, software, governance rules, and workers labeled training data. We refer to such systems using the term "algorithm" as synecdoche [68] and shorthand, since to outsiders, they also appear to be a single, coherent, black-boxed entity. Social media users provide the data, such as textual posts or images, that makes emotion recognition possible, while usually having few ways to influence the operation of these platforms. They have little to no control over data collection and processing practices [145] and future uses of these data, especially by third parties. Ultimately, a common characteristic of contemporary platforms and emotion recognition applications is the unequal power relation between social media users and those profiling them [140,152].

The recent turn towards ever-more surveillance and quantification of affect and emotion suggests a need for critical research on this topic. This paper examines the often-underrepresented perspectives of data subjects on technological development and use, particularly those of social media users. We refer to persons whose data make algorithms such as emotion recognition possible and who are potentially impacted by their outcomes as "data subjects."[1] We understand social media users to be data subjects because engagement with contemporary platforms also involves enrollment in data collection and processing regimes. Throughout this study, we use the term "data subject," as the concept of "the user" has been critiqued for its neglect of power relations on platforms.[2] Scholars have theorized the enactment

[1] We borrow the term "data subject" from scholars like Sarah Igo [83] and Couldry and Mejias [40]. They do not clearly define the term but use it to describe people entangled with data collection and processing technologies. Couldry and Mejias also foreground the normalization of the data subject position, as the lives of ever more people are constantly converted into data streams under data colonialism. They further highlight that people whose data are not explicitly collected are also impacted by increasingly ubiquitous data-driven technologies. The term "data subject" also has a different meaning focused on identifiability in GDPR [1], which is not how it is used in this text.

[2] The concept of "the user" has been critiqued, for instance, in how it frames the relation of humans and computers around usefulness and productivity, thereby hiding socio-technical injustices [99]. It further does not adequality capture the unequal power relations people experience when engaging with and on platforms. People are dependent on platforms' function as communication infrastructure [113] while having little say in their design and data sharing practices. We also considered using the term "affected individual," but ultimately decided on "data subjects" due to the importance of data to platforms and emotion recognition.

of power in the normalization of capturing and appropriating all aspects of human life, including emotions, in the service of capitalism [153] and *data colonialism*[3] [40].

We align our work with data justice-oriented scholarly debates in Human and Computer Interaction (HCI), Social Computing, Science and Technology Studies (STS), and Fairness, Accountability, Transparency and Ethics (FATE) by focusing on the marginalized and excluded voices of data subjects whose data make emotion recognition possible and who are potentially affected by emotion recognition. Our work is specifically concerned with attitudes, expectations, and folk theories on emotion recognition, drawing upon prior research on Computer-Supported Cooperative Work (CSCW) [45,46,55–57,89]. We conceptualize folk theories as non-formal theories laypeople hold to make sense of, explain, and intervene in black-boxed socio-technical systems [46,47]. Such theories "may differ substantially from the institutionalized, professionally legitimated conceptions held by experts and system designers" [55]. In turn, employing folk theories as an analytical framework foregrounds lay data subjects' knowledge of technological assemblages, thereby challenging established expert conceptions on which platform and algorithm designs are often based.

In this work, we are concerned with folk theories on the accuracy of emotion recognition and its implications, highlighting how data subjects explain ascribe high or low accuracy, and resulting risks. We further investigate theories on meaningful transparency for emotion recognition, thereby shedding light on what data subjects deem to be important and enabling knowledge about the technology. We explore the following overarching questions:

> What folk theories do data subjects have about the accuracy and transparency of emotion recognition technologies? What normative expectations do they have for the accuracy and transparency of these technologies? What are the political implications of these folk theories and expectations?

We conducted interviews and used scenarios to examine data subjects' folk theories and normative expectations of accuracy and transparency in emotion recognition technologies (hypothetically) employed on social media. We found that many participants described high accuracy as a source of discomfort. Some even saw it as a threat to their growth and agency, pointing to ambiguity [11,65] and privacy as a desired design principles. Some argued that emotion recognition is accurate based on *Techno-Promise Theories,* or put differently, the belief in the inherent high accuracy of certain technologies such as AI. However, participants also challenged the possibility of accurate emotion recognition technologies. For instance, some questioned whether inner emotions were accessible through technologies and observations *at all.* We further found that participants perceived contemporary transparency practices as insufficient. Consequently, folk theories conceptualizing improved transparency, which we termed *Meaningful Transparency Theories*, centered on emotion recognition as a technological system and its uses in practice. Participants theorized that meaningful transparency would enable them to be more thoughtful about their behavior online. Some also pointed to how transparency would enable them to contest accuracy claims, highlighting the ascribed importance of transparency to verify accuracy claims. Lastly, in conceptualizing our findings through a folk theory lens, we provide five high-level sets of emotion recognition accuracy and transparency folk theories and discuss our findings' implications.

[3] Data colonialism [40] describes a normalization of capitalist exploitation based on captured and processed social data, highlighting how these practices enact power relations reminiscent of histories of colonialism.

2 RELATED WORK

The increasing adoption of emotion recognition is tied to recent advances in big data and the progression of datafication [140]. This technological infrastructure [19] enables what has been termed "data colonialism" [42] and surveillance capitalism [153], a regime that seeks to read, predict, and control people based on big data for profit. In tandem, critical research has emerged concerned with studying digital infrastructure to recenter justice and ethics [39]. Our work contributes to scholarly discourses in these areas centered on what folk theories [68] data subjects employ as conceptual resources and each user's normative expectations about what algorithmic systems *should* be doing. We focus particularly on the algorithmic governance of key dimensions of accuracy and transparency. In the following sections, we highlight insights related to accuracy and transparency from this literature.

2.1 Emotion Recognition Technology

This work is part of an emerging literature in HCI and STS concerned with the critical study of emotion recognition technologies [6,121,128]. These technologies are concerned with inferring emotions based on social media posts [6], videos and images of faces [121], and fitness tracker readings such as data on body temperature and movement [124]. In computer science, technologies surveilling and inferring affective states and emotions first gained increased scholarly attention under the label of "affective computing" in the 1990s [128]. Currently, techniques for detecting and predicting emotions are referred to as Emotional or Emotion AI [6,105]. Prior work has identified problems with these systems, highlighting, for instance, risks and statistical fallacies in deception detection technologies for European border control [121], tensions between public and private interests in the use of emotion recognition in education [105], and potential dehumanization in the context of mental health prediction research [33].

Research on Emotional AI primarily draws on two major theories of emotion [149,151]. The most widely used [29] is the categorical view largely based on Ekman's Basic Emotion Theory [52,53]. It argues for six "universal" emotions: disgust, fear, joy, sadness, anger, and surprise. Another popular notion is the dimensional view, which aims to model emotions as points in a continuous space [151]. These theories are highly contested [149,151] and can also be considered more broadly as organismic models concerned with individual biological states [52,53]. In contrast, the more sociological interactionist view [80,129] posits emotions to be cultural, situated, and communicative instead of biological, discrete, and purely individualistic. Within HCI, "almost immediately" [130] after affective computing emerged, scholars argued against biological state models, pointing to how emotions are "dynamic, culturally mediated, and socially constructed and experienced" [18]. In this work, we do not engage specifically with concepts heavily related to emotion such as affect and mood. Instead, we point to prior work [130] defining these concepts and disentangling their complex relationship with emotion. Our work also highlights how participants partially articulate some of these theories of emotion in their folk theories on emotion recognition. It centers on the often-neglected perspectives of data subjects, whose data enables emotion recognition and who are potentially affected by emotion recognition, often confronted with the most significant risks, and yet not included in technology discourse and development.

2.2 Social Studies of Algorithms

Extant scholarship addresses the social aspects of social media platforms and algorithms [27,67,95]. For example, researchers have investigated critically how journalism may be reshaped through new platform logics and algorithms and what risks this incurs for democratic societies

[21,25,48]. Beyond studying algorithms as impactful black-boxes, scholars have also investigated people's expectations of and attitudes towards algorithms and how they perceive and interact with them [6,16,45,57,118,147]. Prior work has, for instance, explored expectations of contestability in algorithmic content moderation systems and found that social media users desired fundamental systemic changes instead of individual appeals to certain posts [138]. Other research on understandings and expectations around algorithms uses the concept of algorithmic imaginaries [25,144], which communication scholar Taina Bucher defines as "ways of thinking about what algorithms are, what they should be and how they function" [25]. While the concept has significant overlap with "folk theories" and has received considerable attention, in this study, we adopt the terminology of folk theories, developed and adopted in particular within CSCW and CHI [45,46,55,89,116,142].

We understand folk theories as "intuitive, informal theories that individuals develop" [46] to "explain, interpret, and intervene in" [66] black-boxed socio-technical systems [47]. This conceptualization situates theories within the individual and thereby fits our qualitative approach. We do not seek to highlight how general or representative theories are but instead to explore and identify individual theories that aid in understanding the relationship between data subjects and emotion recognition. This definition further centers emotion recognition as a socio-technical system, thereby foregrounding how accuracy and transparency are both constructed as technical properties, but also exist as contested concepts that are debated and co-produced through power relations.

Extensive prior work on folk theories around algorithms on social media [45,46,55–57,89] points to how data subjects make sense of their experiences on platforms while confronted with black-boxed algorithms and how they employ this knowledge to increase their agency and influence within these systems. The definitions employed in the literature vary significantly [47]. For instance, some posit [55] that folk theories are "non-authoritative conceptions of the world that develop among non-professionals and circulate informally." This conception foregrounds the 'circulation' of folk theories in contrast to our definition focused on the individual. Within HCI and CSCW, studies on folk theories provide valuable insights into how laypeople interact with and think about black-boxed algorithms which they encounter regularly but cannot directly understand or shape. They have been employed to rethink established platforms and algorithms and generate policy recommendations [47]. It is important to examine what people think an algorithm does as well as it is to examine what an algorithm actually does precisely because of the impact that folk theories have on people's attitudes and behaviors [137].

2.3 Accuracy

Recent scholarship has problematized accuracy (i.e., number of correct classifications out of all data points) in algorithmic systems as a measure for fairness or closeness to "objective reality" [72]. Scholars pointed out that machine learning algorithms exhibit varying degrees of accuracy for different groups [2,26], e.g., commercial facial recognition products are much less accurate for Black women than white men. Varying degrees of accuracy have been uncovered through algorithmic auditing [123,133]. In the context of emotion recognition technology, researchers found sentiment analysis frameworks to exhibit gendered and racialized intensity biases in emotion detection [92], reproduce occupational gender stereotypes [15], and classify sentences associated with being old as less favorable [49]. Tensions and tradeoffs between accuracy and "group fairness" [106] have received attention in media and academia as disparate impacts [12] were uncovered in widely used applications such as recidivism risk assessment algorithms [10]. Group fairness is most often centered on equal or fair outcomes when comparing different groups using categories protected within US civil rights regulation, such as gender or race: this produces tensions between fairness and accuracy. Credit scoring algorithms optimized for accuracy, for example, recommend limited lending opportunities in majority poor or Black neighborhoods [63,90,134], thereby reinforcing structural inequalities. In response to increasing attention to concerns around algorithmic discrimination, scholars have highlighted how claimed neutrality and objectivity of data [20,70,115] associated with algorithms are fictions that mask internalized social inequalities [2,63] and biases [62]. Critical Race STS scholar Ruha Benjamin [13] has referred to this phenomenon of "coded inequality" normalized through "imagined objectivity" as the "New Jim Code."

Many of these reported findings challenge accuracy as a universal quality measure employed in the computer science literature. However, accuracy as a sense-making concept is also part of broader non-academic discourses. For example, scholars found that trust in the accuracy of algorithms was influenced by accuracy claims and observations of algorithmic behavior [150]. Ultimately, data subjects' ascriptions of accuracy depend on how algorithms are presented and how accurate their behavior is perceived to be. When users associate algorithms with high accuracy, algorithmic results become difficult to challenge. The authority of algorithms was even found to influence some participants to doubt their own judgments about their personality when they disagreed with algorithmically inferred traits [56]. In contrast, a more recent study of emotion recognition in education found that students doubted the system's accuracy and validity and even noted this in a survey without being asked explicitly [141]. These findings highlight how perceptions around the accuracy of algorithms may change over time and can depend on the context and stakes.

Researchers have also considered perceptions of comfort with algorithmic decision-making and found concerns centered on accuracy, such as bias and difficulties in modeling complex realities [23]. Participants in this study also wished for the chance to question the algorithms. Similarly, a student quoted in another study distrusted emotion recognition accuracy and argued for contestability [141]. Another study [147] found that most participants were unaware of issues around algorithmic discrimination, but when informed, their concerns increased. Research participants experienced "algorithm disillusionment" when they learned that algorithmic advertising was far from completely accurate while also preferring incorrect algorithmic assessments when it made them think about themselves in "flattering" ways [56]. In turn, high accuracy has been found to be both desirable and unwanted in certain situations, tensions we explore in the context of emotion recognition. We extend this prior work by specifically focusing on folk theories around the accuracy of algorithms. This focus is important because, as we have highlighted in this review, accuracy is a contested concept with varying understandings and much authority e.g., when people doubt themselves because an algorithm is perceived as more accurate than themselves. Misconceptions around the accuracy of algorithms, especially in such a contested area as emotion recognition, can have problematic consequences and foster self-doubt.

2.4 Transparency

Transparency can aid in making accuracy claims more accountable as it opens possibilities for scrutiny. Within academia, the transparency of algorithmic systems has received attention because of the increasing usage of algorithms deployed as opaque "black boxes" [110]. Scholars have called for a countervisuality of algorithms [122] centered on the interests of those affected by them. Efforts by companies to reveal how algorithmic processes work have omitted certain critical and political information [122,76]. Simple calls for transparency have also been met with critique as information on such systems requires a critical audience with the resources and expertise to hold decision-makers accountable [91]. Scholars have highlighted several other limitations of the transparency ideal [5]: claims of transparency can be disconnected from power, can be harmful, can create false binaries, can invoke neoliberal models of agency, and can privilege seeing over understanding. In practice, technical and temporal limitations constraint transparency. Ultimately, transparency is complex, contested, and multifaceted, and by itself, not sufficient to make algorithmic systems just and democratic. However, it still matters as a necessary condition of procedural justice [16], which is concerned with explaining decision processes and making them visible and accountable to those affected by them.

Perceptions of algorithms regarding transparency, especially concerning explicability [51] and interpretability [16], have received considerable scholarly attention. Researchers have studied the effects of transparency on user behavior and perception by providing explanations for content moderation decisions [87] and ad recommendations [56]. The importance of studying transparency is also made clear through studies highlighting feelings of unease due to its absence. For example, Airbnb hosts [88] believe that a lack of transparency can generate anxiety due to perceived loss of control and knowledge about algorithms on the platform. Prior studies also highlight the complexity of designing for transparency to increase trust and satisfaction [28]. When designing for different audiences, there are tradeoffs about how much information should be shown and in what ways. For example, if not enough information deemed important is provided, concerns and dissatisfaction can arise [96,127]. In contrast, other studies found that too much information can also "erode” trust [96]. Other studies found that explanations and comprehension did not influence trust in algorithmic decision-making processes [34].

This review has highlighted the complexity of both accuracy and transparency and the various ways folk theories, beliefs, and expectations about emotion recognition technologies have been explored in the literature of HCI, STS, and Fairness, Accountability, and Transparency (FAccT). The review describes an active area of research that remains in its early stages, with some conflicting evidence and open questions about the role that folk theories might play in emotional AI. Our work extends this literature by focusing specifically on folk theories, attitudes, and expectations around accuracy and transparency in the context of emotion recognition and highlighting their interrelatedness. Although some previous work has reported insights related to stated beliefs around algorithmic accuracy, ours focuses on data subjects’ folk theories around accuracy and their relationship to transparency. The ubiquity of black-boxed, opaque algorithmic systems highlights difficulties in challenging misconceptions around ascribed high accuracy and its potential authority. Consequently, expectations and folk theories around transparency and accuracy should be analyzed in tandem. In particular, emotion recognition provides a rich context to examine folk theories and expectations of algorithmic accuracy and transparency because emotions have opaque normative weight, are most commonly theorized and modeled in problematized ways, and potentially lead to the reification and internalization of standardized models of emotions [130].

3 METHODS

We conducted a series of 13 remote in-depth semi-structured interviews (77-120 minutes, 106 minutes on average) with adult social media users in the US in the summer of 2019. The IRB approved our study. We deployed a screening survey to purposefully recruit interview participants as described below. Although qualitative in-depth interviews do not rely on a logic of representation for their validity, a purposefully selected sample can improve the quality of the results. To increase diversity along the axes of race and education, we posted the survey link on

Craigslist [148] pages for Detroit and Houston—two of the most diverse cities in the US [103]. We also posted links to the survey on the last author's accounts on public social media such as Twitter and Facebook, which were then shared by many outside the author's network. No interview participants were known to the researcher conducting the interviews. We provided each participant with a $30 honorarium.

Via the screening survey, we required participants to be 18 or older, reside in the US, and use social media. We measured respondent demographics and asked about social media use. To seed the in-depth interviews, we queried respondents about positive and negative life experiences in the past year and asked whether they had shared about these on social media. Participants were encouraged to look at their social media as they responded. The screening survey received 100 responses, but not all were eligible for our study. We invited 20 individuals to participate in interviews; 13 ultimately responded, signed the consent form, and completed an interview.

In deciding whom to invite for interview, we followed an iterative and purposeful process, considering the data, identities, and experiences represented in our collected data at any given time and striving for a diverse range of experiences and identities (along the axes of age, gender, race, and education) to the extent possible. Additionally, our research questions required participants to reflect on their actual past social media sharing behavior about personal, emotional experiences, both positive and negative, to describe how they would feel about emotion recognition based on such data. Therefore, we considered survey respondents who had experienced both positive and negative personal experiences in the past year and posted about them on some social media platform. Those who did not report such experiences were not invited to participate. Examples of participants' positive experiences included getting a job, getting a degree, getting into college, or buying a house. Negative experiences included losing a job, ending a relationship, and mental and physical health challenges.

In line with exploratory and qualitative approaches [14,126], our data was not representative of social media users affected by emotion recognition, and our goal was not representativeness. However, for context, we note the characteristics of our interviewee pool (see Table 1). Our pool skewed young and educated and had a large number of women. Thus, we were able to include perspectives from typically underrepresented genders. We note that men may be less likely to participate in a study about emotions or to share emotional experiences more broadly [22]. We discuss the limitations of our sample in section 3.4. in more detail.

Table 1. Participant demographics. Abbreviations for social media sites: Archive of Our Own: AO3, Discord: DC, Facebook: FB, Facebook Groups: FBG, Instagram: IG, LinkedIn: LI, Reddit: RD, Snapchat: SC, Tumblr: TB, Twitch: TCH, Twitter: TW, YouTube: YT

	Age	Gender	Race	Education	Social Media
P1	24	Agender	White	College	FB, TW, RD, TB
P2	58	Woman	White	Graduate	FB, TW, LI
P3	20	Genderfluid	Indian	College	FB, IG, TW, TB, AO3
P4	23	Woman	Asian	Graduate	FB, IG, TW, RD
P5	25	Woman	White	College	TW, SC, TB, DC
P6	43	Woman	Black	College	FB, FBG, IG
P7	28	Woman	White	Graduate	FB, FBG, IG, TW, SC, RD, LI
P8	36	Woman	White	Graduate	FB, FBG
P9	24	Woman	Asian	Graduate	IG, TW
P10	27	Genderqueer	Black	Graduate	FB, FBG, IG, TW, SC, RD, TCH, YT
P11	22	Man	White	High School	FB, FBG, TW, SC, RD, TB
P12	52	Woman	White	College	FB, FBG, IG
P13	39	Woman	White	Some College	FB, FBG, IG, TW, SC

3.1 Interviews

This paper is part of a larger project about data subjects' attitudes toward emotion recognition technologies. We describe the breadth and depth of the data we collected for the whole project. However, this analysis focuses only on folk theories, attitudes, and expectations about accuracy and transparency. We began the interviews by learning about participants' existing social media use practices; sharing behaviors in relation to personal, meaningful, and emotional experiences; expectations; and understandings of what happens to such data when shared along with expectations for privacy and emotions' meaning. This phase allowed us to understand the context of participants' social media use, especially in relation to personal, emotional experiences. This phase also set the context for the next phase of the interviews by focusing on participants' conceptions of and experiences with emotion-situated experiences. We then relied on scenarios to elicit participants' values, concerns, and attitudes towards emotion recognition on social media. Scenario-like methods are useful tools to elicit values regarding technologies, particularly emerging technologies [4,24,30,77], when people may not have direct experiences to rely on [60]. Responses to scenarios can also help to develop new theory [8]. As far as social media is concerned, companies' lack of transparency about their practices makes it difficult to assess when and how they currently use emotion recognition. However, prior work has highlighted patents and companies active in this space [6,130]. Our study takes social media as one context within which emotion recognition applications are a real possibility (if not an existing reality). It also highlights data subjects' perspectives on two dimensions: accuracy and transparency.

Although a response to a scenario may ultimately differ from subsequent behavior, there is evidence that people tend to react similarly to scenarios in emotional contexts as they would in "reality" [82]. Rather than seeing what people might or might not "do" in the face of emotion recognition technologies, our objective here was to elicit values and attitudes around an emerging technology that is not readily available to non-experts. Informed by prior work on privacy values, folk theories [57,146], and algorithmic imaginaries [25], we take the position that what people think algorithms can do is as important as what they actually do. We also sought to provide flexibility to participants in how they interpreted and imagined the given scenarios.

3.2 Scenarios

Our interlocutors had already reported experiences with sharing positive and negative events in our screening survey. Using scenarios to connect these experiences with possible emotion recognition use allowed us to elicit their folk theories and reactions to emotion recognition in more depth. Scenarios were presented via a link to a Google document. All participants were presented with the same prompts, albeit in different, randomized orders. The text, presented once for positive and once for negative emotional experiences, was as follows:

> *I would like you to think about something [positive/negative and personal] that brought out [positive/negative] emotions for you. Maybe the experiences we talked about earlier. Now consider this scenario: You had shared on [insert social media they use most] about that, and had explicitly shared how you felt about it. Everyone reading it would have been able to understand what your experience was and how you felt, there was no ambiguity. Now imagine that [insert social media they posted on] used computational methods to detect what emotions you felt at the time of posting that.*

The above example is one that considers direct disclosures of emotions; other scenarios included in the appendix were about indirect and non-disclosures of emotions. Computational techniques are developed to infer emotions and emotional states based on direct and explicit pointers to one's state (e.g., "I am sad") that leave no room for interpretation and inferences, those that are more indirect and vague with more room for interpretation, or those that do not involve direct or indirect disclosures [9,102,136]. Thus, the scenarios asked participants to imagine these distinct approaches to algorithmic emotion recognition. We note that these scenarios were intended as speculative [58] as a starting point to gather participants' reactions; differences or similarities between scenarios were not relevant to our RQs here; therefore, the analysis and data do not align responses with prompts, following past best practices (e.g., [8]). In other words, due to the semi-structured nature of the interviews, participants often went back and forth between scenario discussions to make broader points about their expectations of transparency and accuracy. We ensured that the scenarios were clear, understandable, and not confusing by asking several colleagues to vet them. We framed scenarios as neutrally as possible, broadly enough to allow us to probe for topics of interest, and narrowly enough to be understandable to participants and promote rich data.

In all cases, we asked participants to tell us about their imagined experiences and the emotional connections they drew. In all cases, these included emotional experiences they had noted in the screening survey and earlier in the interviews, and sometimes additional past experiences. We thereby established a personal context with the participants and then probed to examine their attitudes, concerns, and reactions towards algorithmic recognition of emotions based on social media data.

We asked them to imagine how they would feel if the social media they posted resulted in computational techniques being used to infer their current emotional states. The interview protocol is available as supplemental material and includes several other topics not relevant to this analysis. We then probed whether, how, and why knowledge of the existence of these emotion recognition-enabled detections and predictions and how they functioned, and how correct or accurate they were mattered in participants' attitudes toward emotion recognition. In these conversations, as we describe in the findings, the themes of accuracy and transparency surfaced. While our interview protocol specified that we would ask about these topics if they did not arise, they surfaced organically in most cases.

We note that these scenarios were not designed or used in an experimental sense to draw connections between various variables but were used as prompts and conversation starters to get at participants' attitudes towards emotion recognition and its dimensions. Future work could use experiments to examine concrete connections between variables.

3.3 Analysis

We analyzed the data using open and axial coding [37]. We began with open coding and engaged in frequent iterative discussions to refine codes and identify patterns through which themes of accuracy and transparency surfaced. Examples of open codes included "wanting to know why information is being collected, wanting transparency in how detections are made, wanting transparency in how detections are used, wanting transparency in what information is known," and "wanting transparency that detections are happening," which when grouped into larger themes, describe aspects of participants' concerns regarding when and what kinds of transparency are desired. A researcher on the team and the second author frequently met during data collection to discuss themes and inform future interviews. The same team member open-coded five interviews. The second author and that team member then discussed each code in detail, refined codes, and grouped them into larger themes. The team member then coded another five interviews and grouped codes into new themes or ones already developed and then coded the remaining interviews (we identified no new themes in this last phase). We stopped further recruitment efforts after our analysis was conducted, as we had been able to surface similar narratives across data sources. The first author then used these codes and themes to address our RQs. They constructed the folk theories based on the themes through interpretation. The author specifically looked for data subjects' explanations of the accuracy of emotion recognition, theories about risks, and theories about what forms of transparency were important. We did not include theories on current transparency practices because participants were unaware of them.

3.4 Limitations

We recruited largely on social media because we wanted to recruit social media users. However, this also meant that some respondents were in our network; we addressed this limitation that could have led to higher self-presentation concerns among participants by ensuring that the person who conducted the interviews was a stranger. That said, self-presentation concerns are an expected limitation in conducting interviews that can be partly reconciled through building rapport and following best practices.

We note that conducting research on emotions may dampen interest of male-identifying people [44] and that this may in part explain the limited response rate from men. Considering the gendered character of emotional expression, future work could specifically investigate differences in attitudes towards emotion recognition in relation to different kinds of emotions and stoicisms. Men's perspectives could also be particularly interesting due to the stigmatization of men's emotional expression [36,81]. Similar to other studies of emerging technologies [3,75], most participants had at least a college degree and may have been technology savvy. Furthermore, college-educated people often also have a higher economic standing, which in turn may dampen their assessment of the riskiness of new surveillance technologies, which usually disproportionally target marginalized people, including those of lower socio-economic status [13,35].

Despite the unique composition of our sample, our study provides valuable insights into perspectives of data subjects on an emerging technology. Our approach, grounded in deep interviews, allows generative and interpretative insights instead of generalizable knowledge. Our goal was not representativeness; additional work in this area could be made stronger by actively including perspectives from less educated people, other genders, diverse races/ethnicities, older adults, children, and non-US citizens. Assessing the prevalence and significance of identified folk theories within the broader population via surveys is an area for future research, as the identified folk theories in this study are to be considered preliminary. As a next step, future work could also seek to research perspectives of people who claim to not be affected by emotion recognition due either to not posting emotional content or not using social media. Due to the ubiquity of social media, they may still be unknowingly affected, for instance, when mentioned in another's post or through algorithmic misclassifications.

We further note that our study used textual scenarios and not *actual* social media interfaces. Understanding the ways interfaces intersect with underlying emotion recognition algorithms, and what that means for expectations of accuracy and transparency, is an area for future work. The scenarios are powerful in allowing us to examine participants' values and beliefs about emotion recognition on a conceptual level; however, they are limited in that they may not necessarily reflect actual behavior associated with such beliefs, though they might. Our study design involved recalling past sharing experiences on social media, and this recall is likely imperfect. That said, because we were interested in folk theories, values, and sense- or meaning-making about emotion recognition, errors in recall would not have affected our investigations.

4 FINDINGS

We first report on participants' stated attitudes and beliefs towards accuracy, then discuss agency in relation to transparency, limitations of transparency, and finally normative expectations of what meaningful transparency looks like in practice. Our review of the literature highlighted how both algorithmic transparency and accuracy are contested, complex, and political concepts in various application domains [5,21,87]. This study does not seek clear-cut definitions of accuracy or meaningful transparency in the context of emotion recognition from participants. The development of such conceptualizations entails further research, broader democratic deliberation, and value-based decisions. The insights we provide into data subjects' beliefs, attitudes, and expectations still aid in governance and design of emotion recognition technologies and reveal problematic assumptions in current approaches. In our analysis, we particularly highlight promises, tensions, and problems around emotion recognition to give insights into the complexities of how emotional recognition technology and its uses in the context of social media are perceived. We emphasize the grave contemporary power imbalances between platforms and data subjects and deficits of contemporary emotion recognition uses on platforms, as stated by participants.

4.1 Accuracy and Agency

Since the static conceptualization of emotions based on a few emotional states [52,53] in many emotion recognition technologies is heavily contested, serious questions on construct validity arise. In turn, it is questionable whether quantifying accuracy is a fruitful endeavor [109] when the underlying modeling assumptions are problematic and faulty. We don't seek to directly address these definitional socio-technical aspects; instead, we are interested in exploring how lay data subjects feel about accuracy in the context of emotion recognition and how they make sense of the technology's capabilities. We find that some perceive highly accurate emotion recognition as uncomfortable, even framing it as a threat to their agency, privacy, and growth. Some participants form expectations based on different kinds of accuracy. For some, discomfort is related to context. Some participants' stated beliefs were also aligned with dominant discourses around promises of big data. However, others brought forth more humanist critiques, which point to competing ideologies about making sense of emotion recognition and, in turn, emotion AI. The following sections unpack these insights.

4.1.1 Stated attitudes towards accurate algorithmic emotion recognition on social media. Some participants perceived accurate emotion recognition as a threat to their agency, while others welcomed it in certain contexts. Some argued that emotion recognition is currently not much of a concern because it is inaccurate, but also pointed to the dangers of undesirably shaping emotions through algorithms and categorization. Ambiguity and inaccuracy were understood by some as empowering and enabling possibilities for agency. Ultimately, we found several stated beliefs and expectations about accuracy, which were also context-dependent.

Some perceived accurate emotion recognition as uncomfortable. For example, P7 said, *"If it's not accurate, I'm probably more okay with it. But if it is accurate, then that feels bizarre."* Inaccurate emotion recognition was assumed to be the status quo by some, who found it is less concerning. P1 said, *"I think computers probably have a hard enough time with it that I'm not that worried about it, not yet. We'll see how good they get at it."* The participant frames less accurate emotion recognition as less problematic.

Some perceived accurate emotion recognition as a threat to their agency and privacy. Commenting on accuracy, P3 said, *"I feel the detection being more accurate would probably freak me out more...like, 'Wow, the AI age is upon us and they're reading all our data. '...It can go wrong very easily or it can be used poorly...if they're not that accurate, I would feel more at ease in terms of the grander scheme of things. But I would still feel really weird about them kind of doing that."* While higher accuracy meant more discomfort to the participant, the use of emotion recognition, to begin with, was perceived as uncomfortable. The feelings were argued to be based on anxieties about risks due to possible "poor" and "wrong" uses of emotion recognition.

Some participants also raised concerns about how accurate emotion recognition could impact a user's emotions, perception of self, or growth. For example, P10 said, *"I think I probably wouldn't like it [emotion recognition], not because it's not useful, but because it can very easily be, like, a self-fulfilling prophecy. Like, if something is trusted and can predict sort of accurately how you're going to feel in the future, then you will think that it's how you will feel, and then you will feel that way, which is not good...especially for folks that are sort of growing up and learning how to feel and navigate feelings, having something tell them how they're going to feel later is not probably beneficial for their growth."* This example warns of emotion recognition being trusted and framed as a technology that accurately infers emotions, since it thereby also becomes a tool to understand and form the self and one's emotions in possibly undesirable ways.

Some participants perceived accurate emotion recognition as practical, but such assessments were dependent on contextual factors. Whether accurate emotion recognition was perceived as beneficial or risky depended on the context of its uses for some. Some perceived accurate recommendations as more genuine to who they were, yet there was still some discomfort with the accuracy of the results. For example, P8 noted that overall, they preferred inferences to be accurate: *"I guess I feel like I would like it if they're getting it right and maybe I wouldn't like it if they're getting it wrong."* But even so, this preference was context-dependent*: "It depends on what they were going to do with it, but like what if they got it wrong. Did I like take a gamble...posting [a] snarky post about my child. Like if they assign intent that didn't really exist then I would be upset I suppose."* P8 continued*: "I mean it does feel contradictory in some way because it's like I want it [the recommendation] when I want it, then I don't want it when I don't want it."* In this case, when and in what context the accurate emotion recognition-enabled recommendation appeared in one's social media feed mattered, but the overall technology was not rejected. Imagined risks of inaccurate inferences and recommendations also shaped participants' attitudes. For instance, if the algorithm suggested that they were a "bad parent," e.g., because the snarky post was understood as a literal feeling about the child, the risks of inaccuracy were seen as significant.

Participants noted that emotion recognition accuracy is not only about correct inferences of emotions but also how to appropriately handle inferences considering the contexts and situations in which they arise. The accuracy of emotion-recognition enabled recommendations (e.g., content, friends) on social media played a role in participants' attitudes. When it came to ads (as one type of recommended content), for some, how relevant an advertisement was affected how comfortable they were with seeing it. For instance, P7 said, *"I mean again, the more accurate it is probably the more okay I might be with it....if I'm planning a wedding and they send me some things that are really helpful, I'm going to be pretty excited about that...if I was sad about a breakup...and then Facebook was like, hey check out this new yoga place and get $10 off, okay...you're tailoring the ads to something that makes sense right now. Versus if I was upset for a breakup and now you're showing me engagement photo photographers, I'd be really bummed."* While P7 noted that some ads could be harmless, they also noted that is not always the case. For example, when ads are insensitive to what one is experiencing or when they take advantage of those experiences: *"If they were using the data to purposely advertise expensive stuff to people that were feeling super vulnerable. I feel like that's harmful."* This statement highlights how the participant expects that emotion recognition should not only accurately detect personally held emotions but also accurately handle the situations in which they arise without inadvertently inducing any harm. It points further to a perceived risk of emotion recognition figuring emotions as individual, decontextualized states, thereby producing unsuitable recommendations that miss contextual nuances. The statement highlights how emotions are situated and in relation to the world. We interpret the participant's framing to suggest further an expectation of different forms of accuracy of emotion recognition systems: one about capturing individual personal states of emotions and the other about understanding and respecting the contexts/situations in which they arise. In the next section, we highlight how some participants also explicitly voiced concerns that emotion recognition cannot capture such complexities.

4.1.2 Stated beliefs towards accurate algorithmic emotion recognition on social media. We identified several stated beliefs of data subjects about emotion recognition's accuracy. Some participants made sense of emotion recognition accuracy based on popular myths about big data and AI. They argued, for instance, that emotion recognition must be accurate due to so much "big" data being available online. Others foregrounded inherent inaccuracies, e.g., because online data is not representative of one's life, or because online posting is performative [41], and does not reflect "true emotions." Finally, some questioned almost entirely the capabilities of emotion recognition in practice.

Based on popular beliefs about big data and AI, some participants were convinced that emotion recognition algorithms would be accurate. For example, P2 said, *"I'm assuming that they're going to be very good at making these predictions because there's just so much data available out there about people's shopping habits and people's posting habits, and what people are doing. There's so much information available that I suspect that they would become fairly good at making [predictions]."* The participant was seemingly enticed by the promises of big data and assumed that all this online data would have to enable emotion recognition technologies to infer emotions accurately. Echoing this sentiment, P7 said, *"I feel like our technology is smart enough to detect that I guess."* In this case, the participant ascribed smartness to emotion recognition and AI, thereby reproducing tropes around artificial intelligence technologies being actually intelligent and therefore accurate.

However, some participants were skeptical of emotion recognition algorithms' ability to accurately infer emotions due to online data being partial, emotions being voiced indirectly, and postings not representing genuine emotions. Some argued that because the data feeding the algorithms (e.g., social media posts) are only about a small portion of a user's life and do not capture everything about them, emotion recognition algorithms would not be accurate. To this point, P6 said, *"You're only looking at...in the grand scheme of things, somebody's overall life, you're only looking at a small portion of that and you're taking that again and grouping them with other people.... They don't know everything about me based on that little bit of information...I guess they can kind of look into what you're saying and kind of get how you're feeling, but at the same time they could also get it wrong."* This statement is an example of a participant reasoning that emotion recognition algorithms would not be accurate because inferences are based on little data. Participants remarked that partial data was not enough to capture the complexity of the individual behind the data and, in turn, could only be inaccurate.

Reflecting on emotion recognition based on pictures, some participants thought their pictures did not accurately reflect their genuine emotions, so any algorithm reading those pictures would generate inaccurate conclusions. As P9 said, *"I think...people brand themselves all the time on social media. So, the picture that I put out then may not...actually [reflect] my emotions."* The participant argues that their private and hidden feelings are invisible to emotion recognition algorithms as their social media content does not reflect their true emotions. Further, P1 questioned the possibility of highly accurate emotion recognition when asked about inferences based on posts that lack any direct and explicit references to their emotions: *"But say I haven't put anything out there, I haven't said how I feel, I implied how I feel, anything like that for a lot or website or whatever to make that prediction feels creepy cause...I wouldn't know how it got there and to make it accurately is even creepier because one, I don't know how about that information and two I don't know how it got that it's accurate."* The participants were skeptical as they could not even imagine how accurate inferences could be achieved in this case.

Additionally, some participants noted that computers and algorithms simply could not understand nuanced human emotions. For instance, P1 said, *"We are literally born and made for being around other people, interacting with other people socially. It's a good survival skill to be able to read another person's emotional state. Even then we get it wrong a lot of the time....often those people who can get it wrong are then building computers that are also imperfect with potentially biased information or inaccurate information....I just don't think that we're ever going to be able to understand or predict emotion computationally like people think we do."* The participant frames emotions as a social practice, which is even hard to grasp and read for humans for whom this task is natural, even necessary for "survival." In contrast to humans, computers were argued to be not intelligent, inferior in their understanding, and built by humans who don't wholly understand emotions.

Similarly, P2 noted, *"I suspect that these sites would probably do a pretty good job at guessing how I would feel about certain things. But they will never capture, or be able to capture, everything about me. The moral of the story being, yeah, statistics can tell us a lot, but they can't tell everything about the individual."* P2 argued that emotion recognition could capture some part but not all of their true emotions. These accounts illustrate that participants believed that algorithms could not accurately read and infer people's emotions because they saw statistics as inherently limited to capturing only certain nuances of the individual.

4.2 Transparency

As mentioned previously, transparency is a contested and political concept [5]. Depending on the context, various ways of enacting transparency can support data subjects and address their concerns. However, transparency practices can also be a form of "Openwashing" [76], concerned mainly with producing a positive public image for platforms through creating the appearance of transparency instead of addressing voiced concerns. In this study, by focusing on data subjects' perspectives on transparency in the context of emotion recognition, we give insights into how they conceive of meaningful transparency. While participants conceptualized meaningful transparency of emotion recognition technology in varying ways, they still overall desired transparency. We first outline how participants imagined transparency would enable them to be more reflexive in their platform use. Then we discuss how participants discussed limitations of transparency. Finally, we analyze what forms of transparency participants desired and discuss how meaningful transparency was described: 1) centered on the technological system; and 2) in uses in practice.

4.2.1 Agency in relation to transparency. Participants argued that various forms of transparency in emotion recognition-enabled systems would impact them in different ways. Many responses were focused on enabling individualized actions through transparency, such as making informed choices and decisions about their behaviors online. Consequently, they argued that current systems lack transparency, which meant that "choices" and "consent" were not meaningful. Participants implied helplessness associated with the status quo. Their reflections illustrated that they did not understand whether and how emotion recognition may be employed on social media. More generally, participants argued that transparency is desirable and a requirement for "fairness," but not a solution to all concerns associated with emotion recognition.

Participants remarked that meaningful transparency would allow them to be more cognizant of their data sharing and disclosure behaviors online. This was argued based on knowing what the emotion recognition results could entail beyond sharing the results with the people they intended to reach. For example, P8 said, *"I like the transparency piece with like how all of us, and me, you're asking about myself, like how can we be a more informed person about where I'm sharing and with whom....Because like I said, I don't understand it and so like helping me understand how that works. So, I think like that would be helpful for transparency....Because then you can make a decision about what you disclose and what you don't disclose in a more informed way."* This call for informed decision-making was linked to notions of consent, as P8 elaborated: *"I think if I truly knew what I was consenting to when I opened Facebook every day it would help me decide whether it's worth it to do it."* This whole statement highlights a status quo characterized by a lack of transparency, which is felt through a missing understanding of how emotion recognition works in practice. The participant views a more consentful design as the solution to this helplessness; data sharing behavior would empower and gives them agency to control emotion recognition and its outcomes.

Transparency did not necessarily entail comfort with emotion recognition. P3 elaborated: *"I mean, I think personally that I would still be uncomfortable, but I probably feel more comfortable on the uncomfortable scale, just because they're telling me exactly what it's doing. Knowledge is power. The more you know, the more you feel like you have more control."* First, the statement illustrates how comfort and discomfort exist on a continuum for this participant. Second, we see how transparency would reduce discomfort and increase trust when the motivations and outcomes of an emotion recognition inference were disclosed, leading to participants feeling more in control.

Participants associated transparency with "fairness" and desired it because they believed it would lead to more "fair" outcomes. The notion of transparency was centered on knowing what was learned about a person based on emotion recognition with the opportunity to understand results. In this sense, P10 argued that if they were given this information, the system would be more "fair." Specifically, P10 said, *"I think it's more fair, and I think it gives people an idea of what's happening, especially if it's in a way that they can understand, just, like, 'This algorithm does these things. We're not going to tell you how it works, but here's the result.'"* The participant thereby also gives insights into how they would imagine meaningful transparency for an emotion recognition algorithm. They are not so much interested in technical details but in descriptions of the processes of an algorithm, specifically what the algorithm does and its produced results. Since emotion recognition technology is seemingly difficult to see and notice, participants even argued that the results used, e.g., for recommending ads or learning about data subjects, should become transparent and visible.

Participants noted they might delete their content or stop using platforms if they were not comfortable with insights gained through transparency. However, without transparency, this choice would not exist for them. As P3 said, *"I feel if they were just outright telling me [about emotion recognition], well then, I would still have a choice to delete my stuff or stop using the platform, but if they don't tell me and they just do it I feel I didn't have a choice"* Similarly, P8 said, *"I think if I had a better understanding of what happened, of what they were doing with my information, I would definitely change my behavior or maybe or I wouldn't but I would at least feel like I was making a real choice. Whereas right now it's sort of like we suffer for like living in ignorant bliss or in a state of disbelief around what is actually happening with my information."* Participants understood transparency as a tool to hold companies accountable if their practices are deemed problematic, e.g., in these cases, by deleting one's account. In turn, they perceived opacity as a factor keeping participants from such measures by upholding and producing ignorance [43], i.e., an absence of knowledge about how and what emotion recognition does.

4.2.2 Opacity and limitations of transparency. Participants argued that the absence of transparency is a source of ignorance. They remarked that it made them unaware of how emotion recognition is used and whether it is in their interest. Furthermore, participants argued that there are inherent limitations to emotion recognition's transparency. They highlighted the difficulty of trusting in transparency practices.

Some were concerned about a lack of transparency, as without information, data subjects would not know whether emotion recognition is accurate. For example, P10 said, *"If you think it's inaccurate, then you think it's inaccurate and you can shrug it off. But if you have no idea how it works and you just assume that smart people made a thing and you believe it, then that's where the problems sort of appear."* This participant highlights that lacking transparency is a source of ignorance as it makes it hard to assess the accuracy of emotion recognition. The proposed form of transparency would enable contestability [78,79,108] and validation of emotion recognition accuracy. Currently, accuracy is ascribed mainly through trust in the expertise of experts and tech companies, as noted by the participant.

Some participants discussed the limitations of transparency. For example, P5 said, *"I would say yes, with the caveat that just because they tell you something doesn't necessarily mean it's the truth. They can tell you things to make you complicit, to make you feel better but without any proof of that, then there's always that little under element of 'Hm. I don't know.' I like physical, tangible proof, I guess, in addition."* The participant argues that even well-intentioned transparency requires trust in emotion recognition companies since provided reports are always black-boxed to some degree, and thereby truthfulness is not entirely verifiable. They further argued for additional "tangible" evidence to strengthen claims to transparency and accountability. The following sections focus on how participants thought transparency could be implemented.

4.2.3 Transparency of the technological system: How does emotion recognition work? The following sections elaborate on how participants imagined meaningful transparency for emotion recognition as a technological system. We identified two themes regarding how participants conceptualize transparency. The first unpacked here focused on how emotion recognition works as a technology. In contrast, the second focused on the uses of emotion recognition and is discussed in the following section. For participants, transparency of the technological system included knowledge about how implicated algorithms work, which data is used, and what knowledge or results are produced about them.

For some, meaningful transparency included knowing "how" the algorithm works in understandable and digestible ways, which on occasion required education or explanations from others. Participants noted the difficulties they face in understanding what happens to their data and that, therefore, transparency should strive to improve understandability to be meaningful. As P9 said, *"I think we just need education and to know what the algorithm means and like how the algorithm works and all of that....I just think that it has to be transparent in the way they manipulate our minds and the way they control our lives and regulate...and control our behaviors and actions."* The participant invokes education as a prerequisite for transparency, highlighting the difficulties of understanding such complex algorithmic systems. Some mentioned having to rely on others to explain algorithms to them if such help was made available. For instance, P5 said, *"You know how something works, even if I'm not a tech person and I don't know anything about tech but if you're willing to tell me how it works, I can always find somebody who does know and understand tech and they can tell me if it's real or not."* This statement further highlights how accessing, seeking out, and comprehending this information is a privilege of its own, relying on one's training or network. Transparency, e.g., in the form of access to technical documentation, is simply not enough for many as they also need to have the expertise and time to understand it [5].

Knowing what data emotion recognition algorithms exactly use was an important dimension of meaningful transparency for some. For example, P9 said, *"We just need to know what [data] exactly are being collected....People need to know what they are doing and how they're collecting data."* As P6 described transparency, *"It's being honest and upfront what information, exactly what information you're going to use...and what research or data is this going to be used for."* Similarly, P4 said, *"I would expect more transparency in what data is being used for what prediction."* It is worth noting that for P6 and P4, transparency also meant clarity regarding what the data would be used for. Similarly, P3 said, *"I would feel the most comfortable if they tell me they're going to do it and they're being transparent. Like but this is what we're doing, we're going to take your emotional data now."* These participants ultimately call for transparency about what data is collected about them and, in turn, used to infer their emotions.

Some participants' notions of meaningful transparency included knowing what information was known about them exactly through applying algorithms to their data—in other words, knowing the emotion recognition algorithms' results. As P1 said, *"Exactly what information about me is known and how it's being used would be more comfortable than receiving the things you get now...like if you get a really hyper-personalized ad or...like very personalized, like weirdly accurate to me."* The participant describes how certain forms of precisely targeted personalization were uncomfortable and suspicious because they felt uncomfortably revealing. They viewed transparency about what data is collected and produced about them and how it is used as an improvement of the current state of affairs. It would enable them to judge how personal the collected or inferred information is and to take appropriate and informed action. It ultimately mattered to the participants how highly accurate personalization was achieved, and in turn, transparency was imagined as a means to check perceived accuracy.

P7 reflected on transparency and noted, *"That [knowing results] would make me feel the most comfortable. That seems kind of awkward to say that, but yeah....Because they're being clear about, okay this is the data we're collecting, this is what we do with it, and this is what we've found. It feels like the most transparent it could be....Yeah, like if it's made explicitly clear it would make me more comfortable."* This notion of transparency points to how results of emotion recognition and technologies that build upon it should be more foregrounded and made visible. Furthermore, transparency regarding what data is collected about individuals is brought forth as a matter of concern, and participants highlight how meaningful transparency involves knowledge about the uses and applications of the data. Overall, participants imagined that meaningful transparency would enable them to know how emotional data is processed, what it is gathered for, and how it is being used. We unpack concerns related to uses in the next section.

4.2.4 Transparency of uses in practice: What is emotion recognition used for? In the previous section, we focused on the transparency of emotion recognition as a technological system. Here, we focus on the second central theme related to how transparency is imagined, namely on the *uses* of emotion recognition. Participants' accounts depict an unsatisfactory status quo and a desire for more transparency about uses. Some argued that applications capable of shaping emotional experiences should adhere to higher transparency standards, and some even argued that emotion recognition uses, in general, should be made public. Ultimately, transparency as an ideal for emotion recognition was held high, as the following quotations highlight.

For some, meaningful transparency included knowing why emotion recognition was done and to what end and for what purpose. As P3 reflected, *"And so I feel like the actual issue would just be like, why are you doing this?"* Similarly, P5 said, *"I guess I would wonder what they wanted the information for. Why do they want to know if I'm happy or not?"* For some, knowing the purpose of emotion recognition could potentially lead to a more welcoming attitude towards the technology's use. As P12 said, *"Again, if they get my permission and they let me know what they're doing ahead of time it's fine. Then I would understand and I would respect it's for their research....You don't feel like your privacy being jeopardized and your information, what are they doing with it? If they tell you ahead of time exactly what it's for, then I would be okay with it."* Transparency in this sense meant that participants were made aware of the purpose of emotion recognition use of their data, which was connected to feelings that their privacy was respected. Some believed that the "reasons" algorithms do something should be public information, differentiated from "how" algorithms work—which they noted could be kept secret. P10 said, *"I think that information should be public. I guess maybe not companies' information. So companies can have their secret algorithm, but it should be public what that algorithm does. Like, what is its purpose? What's the output? Everything in the middle I guess you can hide, because no one can out-money you, no one can sue you to find out. It doesn't matter. But some things I think should just be, like, public information."* This participant argues for the public release of information about what emotion recognition algorithms do and are intended to do.

Some wanted to know what the emotion recognition was used for and whether it was for an imagined "good." As P5 said, *"If it were some sort of study being done to better understand and assist people mentally, especially since the internet seems to be such a hive of toxic interactions, if it's being used to better understand people's brains or some sort of medical or academic level, I could see where that would be fine. But then I would add the caveat that the knowledge that your information is being used for such a purpose is something that you are aware of...guess it determines to what purpose it would be used."* Similarly, P7 stated, *"I would want that information known to me as the user, but also be used for good."* Also, P6 noted, *"First I would like to know what exactly how that could be beneficial. What is it, because that's going to give me something to think about...what could companies be using that information for, that data for to detect if you're happy, if you're sad....What is that being used for?"* Participants discussed questions of "good for whom and what," which resonates with critical humanist perspectives.

5 DISCUSSION

We have highlighted data subjects' conceptions and expectations about emotion recognition (on social media) through the dimensions of accuracy and transparency. This section draws from

STS and HCI literature to first discuss overarching themes and their politics. We discuss how some participants made sense of accuracy in relation to emotion recognition, possibly based on assumptions inherent to dominant discourses around big data and AI. Then we discuss how strong beliefs in the accuracy of emotion recognition were considered by some to be risky, and finally, we illustrate how some challenged the idea of accurate emotion recognition technology. Next, we highlight how participants were concerned about contemporary transparency practices and imagined meaningful alternative transparency. Finally, we elaborate on the relationship between accuracy and transparency. The following subsections also introduce and discuss algorithmic folk theories we derived from the findings through interpretation. The broad categories and corresponding folk theories are listed in Table 2.

Table 2. Five high-level categories of emotion recognition accuracy and transparency folk

Techno-Promise Theories	**Emotion-Shaping Theories**	**Theories of Emotion**	**Technological Limitation Theories**	**Meaningful Transparency Theories**
Big Data	Techno-Deterministic Emotion Shaping	Private Inner Authentic Emotions	Inaccurate Individual Statistics	Transparent Technology Use
Intelligent AI	Emotion Category Reinforcement	Performative Emotions	Unrepresentative Online Data	Transparent Intent
		Interactionist Emotions	Human-Built Technology	Transparent Technological Assemblage
		Emotional Intentions		

5.1 Folk theories of Accurate Emotion Recognition

We encountered participants heuristically making sense of contemporary emotion recognition technologies based on folk theories aligned with popular techno-promises. In Table 2, we refer to these as the *Techno-Promise Theories*. Some participants' statements were seemingly based on assumptions common to big data [93,140,152], referencing the perceived great amount of social data available online as a sign that accurate emotion recognition must already be a reality. We term this the "Big Data" theory. However, various scholars have found that big social media data is insufficient to ensure high accuracy [97,112,135]. In turn, the promise of big data remains tied to a yet-to-arrive future. Therefore, we theorize that big data could also be understood as a socio-technical imaginary[4] [86], a vision of a socio-technical future that is so strong that it is seemingly projected into the present to make sense of the capabilities of contemporary AI technologies such as emotion recognition. Prior work [153,140] highlights the dominance of such big data-based discourses by pointing to dataism [140] as an increasingly popular ideology and how it is also a "commercial" idea [101,153] disseminated and stabilized in great part by companies aiming to sell AI-based products.

Similar to the big data promises, participants in our study also argued that emotion recognition *must* be accurate due to its "smartness" and thereby ascribed intelligence to AI technology. We call this the "Intelligent AI" theory. These ascriptions of intelligence to AI technology [111,120,132] provide a basis for beliefs about emotion recognition accuracy and point to their persuasive power. However, they are also rightly heavily challenged [130]. In this work, we are not mainly concerned with analyzing technology-related discourses and therefore refer to cited prior work. However, this alignment points to possible interesting future research on the power

[4] STS scholar Sheila Jasanoff describes them as "collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through and supportive of advances in science and technology" [86].

relationship of emotion recognition technology discourses and data subjects' folk theories. Within STS, this line of research is usually concerned with investigating socio-technical imaginaries [86,101]. Such work could open possibilities for analysis and critique of what understandings of emotion recognition have been normalized and, in turn, may reify an undesirable status quo, for data subjects, of emotion surveillance. Nevertheless, not all participants' remarks noted a belief that emotion recognition was accurate. Future work could aim to investigate how widespread folk theories based on techno-promises are and, in turn, give insights into the degree to which emotion recognition discourse are captured by simplified and potentially false advertising and promises. Such knowledge could aid in data subject rights advocacy work (e.g., campaigns) and illustrate how widespread problematic conceptions of emotion recognition and other AI technologies are.

5.2 Accuracy as a Risk to Agency

Our findings illustrated a perceived duality of high accuracy in emotion recognition as both a quality to be desired in specific contexts and mostly worrisome and a risk to individual agency. Participants voiced anxieties around undesirable influence on emotions through algorithms and their ascribed accuracy. We discuss these risks later in this section and introduce two related folk theories as part of the *Emotion-Shaping Theories* category (see Table 2). They highlight the dangers of emotion recognition systems prescriptively making their predictions a reality through influencing data subjects. Overall, the participants' statements challenge the notion that ever more accurate emotion recognition is desired by data subjects and, in turn, should be uncritically pursued by researchers and developers in this field. We advocate that both development and scholarship in this space should rather center what data subjects desire instead of pursuing more accuracy. The participants' concerns further point to inaccuracy and ambiguity [11,65] as desired design principles in emotion recognition to protect agency and privacy. This plea also aligns with recent calls in feminist technoscience to see certain glitches and inaccuracies as sources of agency and potentially liberating as they may enable evasion of the algorithmic gaze [119]. Our findings highlight data subjects' desire to only be known by platforms on their own terms when their emotions are involved. The dissatisfaction with the status quo of emotion recognition that pursues accuracy regardless illustrates that data subjects desire more control. Platforms, in turn, should aim to collect and process less data, and when they do, ask for meaningful permission more often. Future research could investigate how desires for ambiguity could be integrated into platforms and what forms of inaccuracy and knowability data subjects desire, which may include completely not being known or seen by platforms.

Some participants worried that emotion recognition perceived as accurate could shape personal emotions to adhere to inferences. They argued that, if believed or normalized as part of platforms, the technology would shape them and the broader society (e.g., through the standardization of emotions), a concern also voiced in previous work [38,115,128]. Ultimately, participants imagined emotion recognition as a technology influencing their emotions and feelings through algorithmically determined categories and coded (human) assumptions of what emotional reactions *should* be. We call this the "Techno-Deterministic Emotion Shaping" theory. Science and technology studies scholars have argued that techno-determinism [145] is a problematic theory since technologies are not inevitable forces that produce social change by themselves. In turn, this folk theory possibly reveals a need for making the social aspects of emotion recognition more visible, such as how data subjects also co-produce the technology with their data or how the technology is based on assumptions about emotions. By foregrounding the humanness and contingency of the technology and including data subjects' perspectives in its conception and governance, worries about the shaping of humans through its ascribed accuracy and power may be lessened. Still, these concerns are important to take seriously and should receive further attention.

Participants also discussed a subvariant of the above-mentioned folk theory. They highlighted the risks of algorithms shaping the development of emotional expressions through the standardization inherent to algorithmic modeling [38,128], reducing the possibilities of individual emotional self-expression. This shaping would create a feedback loop that increases accuracy (as learned emotions adhere more to algorithmic categories), strengthening trust and belief in inferences of emotion recognition [115]. Ultimately, a self-fulfilling prophecy would be enacted, stabilizing emotion recognition categories and inferences while increasing trust in capabilities ascribed to emotion recognition. We call this the "Emotion Category Reinforcement” theory. Scholars have argued that the higher accuracy numbers become, the more their persuasive power increases [61,115]. In turn, this folk theory highlights how creators of emotion recognition technologies need to crucially pay attention to feedback loops and make clear to data subjects what they do to avert such undesired outcomes. It is important to note that these described looping effects [74] between emotion recognition and data subjects need not necessarily lead to an internalization of algorithmic categories but could also result in other (re)actions such as resistance or refusal [64].

5.3 Contesting Accuracy

Beyond intentional harmful misuses and the inference of profoundly personal and private emotions, participants voiced concerns related to limitations of emotion recognition regarding capturing emotions accurately. In this section, we first highlight *Theories of Emotion* that fundamentally question dominant modeling assumptions in emotion recognition. Secondly, we discuss *Technological Limitation Theories*, which challenge the possibility of inferring accurate information about data subjects from online data. Both sets of theories align with various scholarly critiques and illustrate participants’ distrust in the capabilities of emotion recognition. We highlight these critiques throughout this section next to the corresponding folk theories. The theories in this section also illustrate the potential of data subjects to evaluate and possibly co-create assumptions about emotions and technological design necessary in modeling. They further show that data subjects are not uniformly captured by narratives that posit emotion recognition as accurate and capable of capturing genuine emotions. Instead, they show an imaginative and discursive resistance to the supposed inevitability of surveillance capitalism and its promise of total control [153], which is encouraging for a vision of a world that values agency for data subjects. These theories undermine to some degree the powers of the *Emotion-Shaping Theories* presented in the previous section and highlight potential for a broader public debate on the inherent inaccuracies and contingencies of emotion recognition. The critiques also possibly highlight a desire for humility [31,85] in data science and emotion recognition, which asks technology creators not to overpromise on accuracy and to be transparent about contingencies.

Participants argued that their innermost emotions are not capturable by technology as they are private and not visible in online postings. We call this the “Private Inner Authentic Emotions” theory. This observation aligns with previous research on perceptions of “datafication and dataveillance” in which participants argued that AI “could never access their real selves'' [100]. Participants in our study argued further that posting content online to various audiences and publics is performative and not reflective of actual emotional inner states. In this sense, participants’ use of social media platforms to express themselves and their understandings of their performative behavior shaped their perceptions of how accurate emotion recognition could be in practice. We call this the "Performative Emotions" theory. Both folk theories question the possibility of capturing genuine emotion via technological means. They also align with arguments made by various scholars critical of the capabilities of emotion recognition [121,130] and show how lay data subjects resist narratives of accurate emotion recognition that makes visible what may not be visible: inner emotions.

Participants also argued that emotions are social and communicative and cannot be simply captured as categorical emotional data [6], ultimately raising concerns about the construct validity of emotion modeling. This stated view also aligns with an interactionist understanding of emotions [80,129]; thus, we call it the "Interactionist Emotions" theory. Statements of others we also associate with this folk theory pointed to an understanding of emotions as relational and situated in complex contexts. Their understanding was not based on universal, decontextualized, and purely biological states standard in models underlying most of emotion recognition [52,53]. Further, some reframed the optimization goal of emotion recognition beyond capturing "real" emotions. Instead, they called for these technologies to accurately understand the contexts and histories in which emotions arose and to respectfully process emotions. We call this the "Emotional Intentions" theory, which prior work has also identified as a shortcoming of contemporary emotion recognition systems [130].

Beyond the above *Theories of Emotion*, we also identified three *Technological Limitation Theories*. Participants questioned the representativeness of big online social data. They argued that it is only partial and thereby not representative of humans' complex and multiple lives and emotions. We call this the "Unrepresentative Online Data" theory. Such challenges to the representativeness of social media have also been articulated and studied in previous research concerned with big data algorithms [97,112,135] and certainly pose a problem for accuracy. Participants also pointed out how emotion recognition algorithms are created by humans who do not wholly understand emotions themselves. Thereby, computers cannot just "smartly" handle emotions. We refer to this as the "Human-Built Technology" theory. This theory also directly challenges current assumptions in emotion recognition modeling as there is no clear consensus on how emotions and affect should be defined [129], although most emotion recognition technologies draw from Ekman's model that identifies six basic emotions [129]. Others also argued that in their experience statistics are not always accurate when focused on single individuals. Consequently, they also don't expect emotion recognition to be accurate. We call this the "Inaccurate Individual Statistics" theory. Prior work agrees with this characterization, arguing that inner multifaceted (emotional) states cannot be predicted with high accuracy [17].

Overall, all these folk theories show how narratives of accurate emotion recognition are also heavily challenged by data subjects. They illustrate a need within the emotional AI community to rethink and debate current practices based on theories of emotion contested by scholars [42,130] and data subjects. In particular, the possibility of capturing genuine inner emotions is heavily questioned. Consequently, emotion recognition could adapt its presentation and design inspired by the presented folk theories. For instance, companies could seek to communicate the limitations of the technology to capture actual felt emotions and instead present it as a political and potentially risky technology that classifies patterns of emotional performances that require contextual and personal information for further interpretation. Ultimately, the definitional tensions illustrate that emotion AI and recognition deserve more regulatory and scholarly attention as current modeling assumptions don't reflect a democratically deliberated understanding of emotion yet potentially affect data subjects' lives in important ways. Emotions are valued, personal, and political, and technologies that read and process them should strive for democratic participation instead of imposing assumptions about what emotions are and how they should be handled.

5.4 Meaningful Transparency

The participants overwhelmingly expressed feelings of helplessness, uncertainty, and unknowability towards emotion recognition technology and possibilities to shape its usage. All argued that contemporary transparency practices are lacking, with some also citing, without being asked, a complete unawareness of the technology. We Identified three broad folk theories regarding what kinds of transparency participants desired, which we refer to as *Meaningful Transparency Theories.* The purpose of these theories is more consentful platform-design since the strategies data subjects articulated (further explored in section 5.5.) were mainly concerned with enabling individual informed action. Prior work seeks to meet these desires, for instance, by exploring how to rethink platform design centering consent [84] or how to better inform social media users about the privacy risks of individual postings [143]. Participants argued for transparency about 1) where/when emotion recognition is used, and 2) for what purposes, e.g., in the interest of those affected by emotion recognition or a cause they perceive as a "good." We call this the "Transparent Technology Use" theory. It aligns with recent work on consent in social media research based on scraped content, which concluded that participants wanted to know why their data was collected and wanted to be asked for meaningful consent [59]. This form of transparency can be implemented more thoroughly, but it remains challenging to assess when and how often data subjects should be asked and in what form. This is an area for future work.

The interviewed participants were also interested in social media companies' reasons for using emotion recognition. We call this the "Transparent Intent" theory. This theory presumes that intention is non-trivially determinable. However, there can be many multi-faceted reasons for emotion recognition to be employed by social media companies. In turn, finding the right level of abstraction and ways to explain involves difficult and political choices [114]. Furthermore, there may be intentions that companies seek to hide because they conflict with the interests of data subjects, which further complicates implementing transparency. Beyond emotion recognition's uses, participants were interested in the transparency of the technological system itself, which included the data, the algorithm, and produced results. We call this the "Transparent Technological Assemblage" theory. This form of transparency is non-trivial in terms of how to bound the technological assemblage to describe and to do it in an understandable and emancipatory way. As we highlight in section 5.5., this is likely too much effort for most individuals, and in turn, we argue that collective approaches would need to be imagined. Workers involved in emotion recognition assemblage were not brought up by participants organically or in response to prompts (e.g., "who do you think sees your data?"). This is concerning considering that previous research has also highlighted the exploitative nature of certain tasks [69,71] (e.g., data labeling) that enable algorithms. We advocate that any meaningful transparency approaches should also include issues around supply chains, such as independent assessments of working conditions and climate impacts [71,73,107].

Generally, the participants' responses were quite vague (by technical standards) regarding how to implement meaningful transparency, but this is not surprising as they were not familiar with the technical details of emotion recognition. The implementation of transparency in practice is challenging and requires reflexive deliberation with various people [50,98]. It also is not always in the interest of powerful actors since, e.g., components of algorithmic systems [68] are often trade secrets [110]. Additionally, it is difficult to define what meaningful transparency should entail to enable the scrutinization of emotion recognition accuracy. In particular, emotions are dynamic and cultural, not easily enclosed in definable categories [129]. They are already hard to articulate and assess for researchers, and for data subjects, this may be even more challenging. Even if some form of meaningful transparency can be established, many political and operational challenges such as power imbalances and inclusion of data subjects within decision-making processes remain. Still, our findings highlight areas in which data subjects argue that transparency should be improved and point to how transparency is currently seemingly mostly understood in abstract terms by data subjects. Future research could aim to co-develop concrete transparency practices or prototypes beneficial to data subjects that work with their own understandings and folk theories. However, as we discuss in the next section, it is also important to enable collective accountability.

5.5 Contested Transparency Expectations

Our analysis shows that most participants envisioned transparency as a means to improve unsatisfactory implementations of emotion recognition by enabling individualistic notions of accountability. For example, participants explained how they imagined meaningful transparency would enable them to be more reflexive about their content sharing behavior and leave platforms when they disagreed with certain practices. These counter actions are not collective; they aim to elicit change through individual action. Ultimately, most participants understood transparency mostly as a source of individual agency. A few argued that transparency could enable the formation of critical audiences [91] that interrogate the technology, its uses, and how they align with participants' interests. However, no statement was concerned with building up and organizing such an audience or collective. Prior research has highlighted how central critical audiences [91] are to enacting accountability within algorithmic systems. Individualistic conceptions of accountability are limited in their impact as big companies and institutions have more power and can ignore individual acts of resistance. Furthermore, individual interventions are not equally available to all data subjects, especially since platform use is often tied to social support and information networks [7,54]. For instance, leaving the platform could incur considerable costs for some data subjects and further exacerbate inequalities. Also, retrieving, interpretating, and assessing information provided through transparency practices requires knowledge, expertise, and time of data subjects to enable and justify possible individual actions. Some participants also raised such concerns. One argued that any form of transparency still requires trust in the organizations that provide it, highlighting that transparency is not a simple fix for power imbalances.

Scholars have argued that while social media companies have started to communicate more about how their algorithms work, these often efforts fall short of enabling meaningful critique and accountability [76,122]. The current efforts could be understood as public relations work aiming to influence discourse about how algorithms work to create the impression that transparency practices are adequate and algorithms accurate. We seek with this work to intervene in this discursive capturing of how transparency and accuracy of emotion recognition are understood. We aim to do this by centering and discussing perspectives of data subjects at a time when the technology is not completely normalized and stabilized [6], which means alternative trajectories can still be imagined and implemented. It will likely also require collective efforts by data subjects to build up a powerful voice that can shape the future of emotion recognition and imagine accountability beyond the individual.

5.6 On the Relation Between Transparency and Accuracy

We have highlighted accuracy and transparency as complex and contested concepts that hold discursive power. It matters how they are framed and understood as they shape trust and affect towards platforms and emotion recognition, which stabilize them. Participants in our study pointed to a connection between transparency and accuracy. They argued that meaningful transparency enables validation and contestability of results and accuracy. Some further argued that opacity would be a source of ignorance as it impedes verifying emotion recognition results. Ultimately, transparency and accuracy must be considered together in platform design since how the accuracy of a technology is perceived matters, and transparency plays a big part in that.

Prior work [38,56,150] indicates that some data subjects have such a strong belief in certain algorithms that they even question their self-image. Relatedly, various participants also voiced fears about their emotions and societies being shaped by emotion recognition. These insights highlight that accuracy ascriptions and perceptions are important to equitable algorithm design and need to be considered within transparency practices. Furthermore, assumptions about accuracy as a property of algorithms matter. For instance, several participants described emotion recognition's accuracy as a singular factor, but as research into biases of statistical systems has shown, accuracy measures vary depending on contextual and class-/population-level factors. For example, facial recognition software identifies Black women less accurately compared to white counterparts [26]. In turn, equitable transparency and accountability efforts also need to challenge such problematic assumptions about algorithmic accuracy and point to how inequality, power, and history also shape accuracy and how and for whom such performance measures are evaluated. One approach that could aid such efforts is algorithmic audits [123], which enable inquiries into biases that disadvantage marginalized groups and thereby move beyond the individual contesting incorrect or problematic results.

Our discussion further illustrated fundamental conceptual difficulties involved in computationally capturing emotions (as described in section 5.3). Increasing transparency may further reveal some of these issues, e.g., algorithmic misinterpretations, and therefore may be undesired by companies, but making these contingencies clearer can also increase the comfort and trust of data subjects in the long term. It also may enable data subjects to rectify inferences when they want to be better known by platforms, thereby increasing overall accuracy of recognition. Aligned with recent efforts in meaningful contestability implementation in content moderation [138], future work could investigate how emotion recognition algorithms could be contested by data subjects both individually and collectively. The current opaque practices will likely fuel further distrust and may ultimately challenge the technology due to collective frustration. Our findings highlight discomfort with current emotion recognition practices and point to a dire need to center data subjects' concerns in the development, research, and regulation of emotion recognition.

6 CONCLUSIONS

This study highlights folk theories, attitudes, and expectations of accuracy and transparency in hypothetical emotion recognition technologies employed in social media. Our analysis points to an unsatisfactory status quo for data subjects shaped by power imbalances and a lack of reflexivity and democratic deliberation within platform governance. Some folk theories are seemingly grounded in dominant techno-promises and assume emotion recognition to be accurate. Others question fundamentally whether emotion recognition can work at all. Whether data subjects understand emotion recognition as accurate matters since such perceptions influence how and the extent to which the technology is adopted and believed. Many described, in turn, the algorithmic shaping of emotions as a concerning risk and high accuracy as uncomfortable, even a threat to agency. These insights can aid in rethinking the current drive for ever more accuracy and how to discuss and present emotion recognition and its contingencies. Similarly, the folk theories on meaningful transparency could aid in improving design to better center consent. However, as mentioned in the discussion, it is important to institutionalize collective accountability mechanisms, which were concerningly not mentioned by the interviewed data subjects.

Our study points ultimately to a need for intervention to further center data subjects' perspectives. Those who create and deploy emotion recognition must critically reflect on the technology and the fundamental challenges mentioned by data subjects. Our interviews highlight a significant unease with contemporary practices. These feelings could result in more counter actions and dissent being voiced in the future (e.g., platform boycott as mentioned by some participants). An emancipatory understanding of accuracy, justice, accountability, and transparency both as discursive concepts and socio-technical properties in the context of emotion recognition are pressing political and democratic questions that require attention. We contribute to these conversations and highlight folk theories, attitudes, and expectations regarding important dimensions of emotion recognition technologies, hoping to open up possibilities for critical conversations and further research.

ACKNOWLEDGMENTS

We are thankful to the participants who generously shared their perspectives with us, Justin Buss for assisting in data collection and initial phases of the project during his UMSI REMS fellowship, Silvia Lindtner, Christian Sandvig and Cindy Lin for comments on earlier drafts, and the anonymous reviewers and associate chairs for their constructive and encouraging feedback. We also thank the National Science Foundation (award number 2020872) for supporting this project.

A SCENARIOS

Scenario 1:

You had shared on [insert social media they use most] about that, and had explicitly shared how you felt about it. Everyone reading it would have been able to understand what your experience was and how you felt. Now imagine that [insert social media they posted on] used computational methods to detect what emotions you felt at the time of posting that.

Scenario 2:

You had hinted to that on [insert social media they use most], and very vaguely shared how you felt about it. Not everyone reading it, or perhaps no one reading it, would have been able to understand what your experience actually was and how you felt. But you knew what you were talking about. Now imagine that [insert social media they posted on] used computational methods to detect what emotions you felt at the time of posting that, even though you never explicitly wrote anything.

Scenario 3:

You had not shared on [insert social media they use most] about X – this means you have not explicitly or vaguely shared how you felt about it. But you may have done other things online, such as shopping or seeking information or reading content about X or even about other things.

REFERENCES

[1] Art. 4 GDPR – Definitions. *General Data Protection Regulation (GDPR).* Retrieved January 12, 2021 from https://gdpr-info.eu/art-4-gdpr/

[2] Doris Allhutter, Florian Cech, Fabian Fischer, Gabriel Grill, and Astrid Mager. 2020. Algorithmic profiling of job seekers in Austria: how austerity politics are made effective. *Frontiers in Big Data* 3. https://doi.org/10.3389/fdata.2020.00005

[3] Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Transactions on Computer-Human Interaction* 26, 3: 17:1-17:28. https://doi.org/10.1145/3311956

[4] Tawfiq Ammari, Sarita Yardi Schoenebeck, and Meredith Ringel Morris. 2014. Accessing social support and overcoming judgment on social media among parents of children with special needs. In *Eighth International AAAI Conference on Weblogs and Social Media.*

[5] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20, 3: 973–989.

[6] 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), 1–16. https://doi.org/10.1145/3313831.3376680

[7] Nazanin Andalibi and Andrea Forte. 2018. Announcing pregnancy loss on Facebook: A decision-making framework for stigmatized disclosures on identified social network sites. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14.

[8] Nazanin Andalibi and Andrea Forte. 2018. Responding to Sensitive Disclosures on Social Media: A Decision-Making Framework. *ACM Transactions on Computer-Human Interaction* 25, 6: 31:1-31:29. https://doi.org/10.1145/3241044

[9] Nazanin Andalibi, Margaret E Morris, and Andrea Forte. 2018. Testing waters, sending clues: Indirect disclosures of socially stigmatized experiences on social media. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW: 1–23.

[10] Julia Angwin and Jeff Larson. 2016. Machine Bias. *ProPublica.* Retrieved October 29, 2018 from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[11] Paul M. Aoki and Allison Woodruff. 2005. Making space for stories: ambiguity in the design of personal communication systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '05), 181–190. https://doi.org/10.1145/1054972.1054998

[12] Solon Barocas and Andrew D. Selbst. 2016. Big data's disparate impact. *Cal. L. Rev.* 104: 671.

[13] Ruha Benjamin. 2019. *Race after technology: Abolitionist tools for the new jim code.* John Wiley & Sons.

[14] H. Russell Bernard and Harvey Russell Bernard. 2012. *Social research methods: Qualitative and quantitative approaches.* Sage.

[15] Jayadev Bhaskaran and Isha Bhallamudi. 2019. Good Secretaries, Bad Truck Drivers? Occupational Gender Stereotypes in Sentiment Analysis. *arXiv:1906.10256 [cs].* Retrieved August 14, 2020 from http://arxiv.org/abs/1906.10256

[16] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. "It's Reducing a Human Being to a Percentage": Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 1–14. https://doi.org/10.1145/3173574.3173951

[17] Abeba Birhane. 2021. The Impossibility of Automating Ambiguity. *Artificial Life*: 1–18. https://doi.org/10.1162/artl_a_00336

[18] Kirsten Boehner, Rogério DePaula, Paul Dourish, and Phoebe Sengers. 2007. How emotion is made and measured. *International Journal of Human-Computer Studies* 65, 4: 275–291.

[19] Geoffrey C. Bowker and Susan Leigh Star. 1999. *Sorting things out : classification and its consequences.* MIT Press.

[20] danah boyd and Kate Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* 15, 5: 662–679.

[21] David R. Brake. 2017. The Invisible Hand of the Unaccountable Algorithm: How Google, Facebook and Other Tech Companies Are Changing Journalism. In *Digital Technology and Journalism: An International Comparative Perspective*, Jingrong Tong and Shih-Hung Lo (eds.). Springer International Publishing, Cham, 25–46. https://doi.org/10.1007/978-3-319-55026-8_2

[22] Robert Brannon. 1976. The male sex role and what its done for us lately. *The Forty-Nine Percent Majority, edited by R. Brannon and D. David. Reading, MA: AddisonWesley*: 145.

[23] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-making in Child Welfare Services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–12. https://doi.org/10.1145/3290605.3300271

[24] Jed R. Brubaker, Lynn S. Dombrowski, Anita M. Gilbert, Nafiri Kusumakaulika, and Gillian R. Hayes. 2014. Stewarding a legacy: responsibilities and relationships in the management of post-mortem data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 4157–4166.

[25] Taina Bucher. 2017. The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society* 20, 1: 30–44. https://doi.org/10.1080/1369118X.2016.1154086

[26] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*, 77–91.

[27] Jenna Burrell, Zoe Kahn, Anne Jonas, and Daniel Griffin. 2019. When Users Control the Algorithms: Values Expressed in Practices on Twitter. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW: 138:1-138:20. https://doi.org/10.1145/3359240

[28] Carrie J. Cai, Jonas Jongejan, and Jess Holbrook. 2019. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (IUI '19), 258–262. https://doi.org/10.1145/3301275.3302289

[29] Rafael A. Calvo and Sidney D'Mello. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing* 1, 1: 18–37.

[30] John M. Carrol. 1999. Five reasons for scenario-based design. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, 11-pp.

[31] John Carson. 2020. Quantification–Affordances and Limits. *Scholarly Assessment Reports* 2, 1.

[32] Daniel Carter. 2018. Reimagining the Big Data assemblage. *Big Data & Society* 5, 2: 2053951718818194. https://doi.org/10.1177/2053951718818194

[33] Stevie Chancellor, Eric P. S. Baumer, and Munmun De Choudhury. 2019. Who is the "Human" in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW: 147:1-147:32. https://doi.org/10.1145/3359249

[34] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'Connell, Terrance Gray, F. Maxwell Harper, and Haiyi Zhu. 2019. Explaining Decision-Making Algorithms through UI: Strategies to Help Non-Expert Stakeholders. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 1–12. https://doi.org/10.1145/3290605.3300789

[35] Sami Coll. 2014. Power, knowledge, and the subjects of privacy: understanding privacy as the ally of surveillance. *Information, Communication & Society* 17, 10: 1250–1263.

[36] R. W. Connell. 2005. *Masculinities.* University of California Press, Berkeley, Calif.

[37] Juliet Corbin and Anselm Strauss. 2014. *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Sage publications.

[38] Dan Cosley, Shyong K. Lam, Istvan Albert, Joseph A. Konstan, and John Riedl. 2003. Is seeing believing?: how recommender system interfaces affect users' opinions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 585–592.

[39] Sasha Costanza-Chock. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need.* MIT Press.

[40] Nick Couldry and Ulises A. Mejias. 2019. *The costs of connection: How data is colonizing human life and appropriating it for capitalism.* Stanford University Press.

[41] Rob Cover. 2012. Performing and undoing identity online: Social networking, identity theories and the incompatibility of online profiles and friendship regimes: *Convergence.* https://doi.org/10.1177/1354856511433684

[42] Kate Crawford. 2021. *The atlas of AI.* Yale University Press.

[43] Jennifer L. Croissant. 2014. Agnotology: Ignorance and absence or towards a sociology of things that aren't there. *Social Epistemology* 28, 1: 4–25.

[44] Deborah Sarah David and Robert Brannon. 1976. *The Forty-nine percent majority: The male sex role.* Addison-Wesley Pub. Co, Reading, Mass.

[45] Michael A. DeVito, Jeremy Birnholtz, Jeffery T. Hancock, Megan French, and Sunny Liu. 2018. How People Form Folk Theories of Social Media Feeds and What It Means for How We Study Self-Presentation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 120.

[46] Michael A. DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. "Algorithms ruin everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 3163–3174. https://doi.org/10.1145/3025453.3025659

[47] Michael A. DeVito, Jeffrey T. Hancock, Megan French, Jeremy Birnholtz, Judd Antin, Karrie Karahalios, Stephanie Tong, and Irina Shklovski. 2018. The Algorithm and the User: How Can HCI Use Lay Understandings of Algorithmic Systems? In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3170427.3186320

[48] Nicholas Diakopoulos. 2015. Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism* 3, 3: 398–415.

[49] Mark Diaz, Isaac Johnson, Amanda Lazar, Anne Marie Piper, and Darren Gergle. 2018. Addressing Age-Related Bias in Sentiment Analysis. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* - CHI '18, 1–14. https://doi.org/10.1145/3173574.3173986

[50] Lilian Edwards and Michael Veale. 2017. Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for. *Duke Law & Technology Review* 16: 18.

[51] Malin Eiband, Sarah Theres Völkel, Daniel Buschek, Sophia Cook, and Heinrich Hussmann. 2019. When people and algorithms meet: user-reported problems in intelligent everyday applications. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (IUI '19), 96–106. https://doi.org/10.1145/3301275.3302262

[52] Paul Ekman. 2004. Emotions revealed. *Bmj* 328, Suppl S5.

[53] Paul Ekman and Wallace V. Friesen. 2003. *Unmasking the face: A guide to recognizing emotions from facial clues.* Ishk.

[54] Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. 2007. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of computer-mediated communication* 12, 4: 1143–1168.

[55] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First i like it, then i hide it: Folk theories of social feeds. In *Proceedings of the 2016 cHI conference on human factors in computing systems*, 2371–2382.

[56] Motahhare Eslami, Sneha R. Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. 2018. Communicating Algorithmic Process in Online Behavioral Advertising. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 1–13. https://doi.org/10.1145/3173574.3174006

[57] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. " I always assumed that I wasn't really that close to [her]" Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 153–162.

[58] Casey Fiesler. 2021. Innovating like an optimist, preparing like a pessimist: Ethical speculation and the legal imagination. *Colorado Technology Law Journal* 19, 1.

[59] Casey Fiesler and Nicholas Proferes. 2018. "Participant" Perceptions of Twitter Research Ethics. *Social Media+ Society* 4, 1: 2056305118763366.

[60] Janet Finch. 1987. The vignette technique in survey research. *Sociology* 21, 1: 105–114.

[61] Fabian Fischer. 2019. The Accuracy Paradox of Algorithmic Classification. In *Conference Proceedings of the 18th Annual STS Conference Graz 2019: Critical Issues in Science, Technology and Society Studies*, 105-120.

[62] Batya Friedman and Helen Nissenbaum. 1996. Bias in computer systems. *ACM Transactions on Information Systems* (TOIS) 14, 3: 330–347.

[63] Oscar H. Gandy. 2016. *Coming to terms with chance: Engaging rational discrimination and cumulative disadvantage.* Routledge.

[64] Patricia Garcia, Tonia Sutherland, Marika Cifor, Anita Say Chan, Lauren Klein, Catherine D'Ignazio, and Niloufar Salehi. 2020. No: Critical Refusal as Feminist Data Practice. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing* (CSCW '20 Companion), 199–202. https://doi.org/10.1145/3406865.3419014

[65] William W. Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03), 233–240. https://doi.org/10.1145/642611.642653

[66] Susan A Gelman and Cristine H Legare. 2011. Concepts and folk theories. *Annual review of anthropology* 40: 379–398.

[67] Tarleton Gillespie. 2012. Can an algorithm be wrong? *Limn* 1, 2: 9.

[68] Tarleton Gillespie. 2016. Algorithm. In *Digital Keywords* (edited by Ben Peters). Princeton, N.J.: Princeton University Press.

[69] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media.* Yale University Press.

[70] Lisa Gitelman. 2013. *Raw data is an oxymoron.* MIT press.

[71] Mary L. Gray and Siddharth Suri. 2019. *Ghost work: how to stop Silicon Valley from building a new global underclass.* Houghton Mifflin Harcourt, Boston.

[72] Ben Green. 2020. The False Promise of Risk Assessments: Epistemic Reform and the Limits of Fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (FAT*'20). ACM. https://doi.org/10.1145/3351095.3372869.

[73] Gabriel Grill. 2021. Future protest made risky: Examining social media based civil unrest prediction research and products. *Computer Supported Cooperative Work* (CSCW): 1-29.

[74] Ian Hacking. 1995. The looping effects of human kinds. In *Causal cognition: A multidisciplinary debate.* Clarendon Press/Oxford University Press, New York, NY, US, 351–394.

[75] Foad Hamidi, Morgan Klaus Scheuerman, and Stacy M. Branham. 2018. Gender Recognition or Gender Reductionism? The Social Implications of Embedded Gender Recognition Systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 1–13. https://doi.org/10.1145/3173574.3173582

[76] Maximilian Heimstädt. 2017. Openwashing: A decoupling perspective on organizational transparency. *Technological forecasting and social change* 125: 77–86.

[77] Andrew C. High, Anne Oeldorf-Hirsch, and Saraswathi Bellur. 2014. Misery rarely gets company: The influence of emotional bandwidth on supportive communication on Facebook. *Computers in Human Behavior* 34: 79–88.

[78] Mireille Hildebrandt. 2017. *Privacy As Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning*. Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.3081776

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

[80] Arlie Russell Hochschild. 2012. *The managed heart: Commercialization of human feeling*. Univ of California Press.

[81] Bell Hooks. 2004. *The will to change: Men, masculinity, and love*. Beyond Words/Atria Books.

[82] 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: 381–400.

[83] Sarah E. Igo. 2018. *The known citizen: A history of privacy in modern America*. Harvard University Press.

[84] Jane Im, Jill Dimond, Melody Berton, Una Lee, Katherine Mustelier, Mark S. Ackerman, and Eric Gilbert. 2021. Yes: Affirmative Consent as a Theoretical Framework for Understanding and Imagining Social Platforms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery*, New York, NY, USA, 1–18. http://doi.org/10.1145/3411764.3445778

[85] Sheila Jasanoff. 2005. Technologies of humility: Citizen participation in governing science. In *Wozu Experten?* Springer, 370–389.

[86] Sheila Jasanoff and Sang-Hyun Kim. 2015. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226276663.001.0001

[87] Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2019. Does Transparency in Moderation Really Matter? User Behavior After Content Removal Explanations on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW: 150:1-150:27. https://doi.org/10.1145/3359252

[88] Shagun Jhaver, Yoni Karpfen, and Judd Antin. 2018. Algorithmic Anxiety and Coping Strategies of Airbnb Hosts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 1–12. https://doi.org/10.1145/3173574.3173995

[89] Nadia Karizat, Daniel Delmonaco, Motahhare Eslami, and Nazanin Andalibi. 2021. Algorithmic Folk Theories and Identity: How TikTok Users Co-Produce Knowledge of Identity and Engage in Algorithmic Resistance. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2: 1–44.

[90] Michael Kearns and Aaron Roth. 2019. *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.

[91] Jakko Kemper and Daan Kolkman. 2018. Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*: 1–16.

[92] Svetlana Kiritchenko and Saif M. Mohammad. 2018. Examining gender and race bias in two hundred sentiment analysis systems. *arXiv preprint arXiv:1805.04508*.

[93] Rob Kitchin. 2014. The real-time city? Big data and smart urbanism. *GeoJournal* 79, 1: 1–14. https://doi.org/10.1007/s10708-013-9516-8

[94] Rob Kitchin and Tracey Lauriault. 2014. *Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work*. Social Science Research Network, Rochester, NY. Retrieved June 12, 2020 from https://papers.ssrn.com/abstract=2474112

[95] James Kite, Bridget C. Foley, Anne C. Grunseit, and Becky Freeman. 2016. Please Like Me: Facebook and Public Health Communication. *PLOS ONE* 11, 9: e0162765. https://doi.org/10.1371/journal.pone.0162765

[96] René F. Kizilcec. 2016. How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2390–2395.

[97] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. The parable of Google Flu: traps in big data analysis. *Science* 343, 6176: 1203–1205.

[98] Min Kyung Lee, Anuraag Jain, Hea Jin Cha, Shashank Ojha, and Daniel Kusbit. 2019. Procedural Justice in Algorithmic Fairness: Leveraging Transparency and Outcome Control for Fair Algorithmic Mediation. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW: 182:1-182:26. https://doi.org/10.1145/3359284

[99] Cindy Lin and Silvia Margot Lindtner. 2021. Techniques of Use: Confronting Value Systems of Productivity, Progress, and Usefulness in Computing and Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery*, New York, NY, USA, 1–16. http://doi.org/10.1145/3411764.3445237

[100] Deborah Lupton. 2020. Thinking With Care About Personal Data Profiling: A More-Than-Human Approach. *International Journal of Communication* 14, 0: 19.

[101] Astrid Mager and Christian Katzenbach. 2021. Future imaginaries in the making and governing of digital technology: Multiple, Contested, Commodified. *New Media & Society* 23, 2: 223–236. https://doi.org/10.1177/1461444820929321

[102] Lydia Manikonda and Munmun De Choudhury. 2017. Modeling and Understanding Visual Attributes of Mental Health Disclosures in Social Media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery*, New York, NY, USA, 170–181. Retrieved July 9, 2021 from http://doi.org/10.1145/3025453.3025932

[103] Adam McCann. 2019. Most Diverse Cities in the U.S. *WalletHub*. Retrieved September 7, 2020 from https://web.archive.org/web/20190923085620/https://wallethub.com/edu/most-diverse-cities/12690/

[104] Andrew McStay. 2016. Empathic media and advertising: Industry, policy, legal and citizen perspectives (the case for intimacy). *Big Data & Society* 3, 2: 205395171666686. https://doi.org/10.1177/2053951716666868

[105] Andrew McStay. 2019. Emotional AI and EdTech: serving the public good? *Learning, Media and Technology* 0, 0: 1–14. https://doi.org/10.1080/17439884.2020.1686016

[106] Marius Miron, Songül Tolan, Emilia Gómez, and Carlos Castillo. 2020. Evaluating causes of algorithmic bias in juvenile criminal recidivism. *Artificial Intelligence and Law*. https://doi.org/10.1007/s10506-020-09268-y

[107] Thomas S. Mullaney, Benjamin Peters, Mar Hicks, and Kavita Philip (eds.). 2020. *Your computer is on fire*. The MIT Press, Cambridge, Massachusetts.

[108] Deirdre K. Mulligan, Daniel Kluttz, and Nitin Kohli. 2019. Shaping Our Tools: Contestability as a Means to Promote Responsible Algorithmic Decision Making in the Professions. *Available at SSRN 3311894*.

[109] Abigail Z. Jacobs and Hanna Wallach. 2021. Measurement and fairness. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 375–385.

[110] Frank Pasquale. 2015. *The black box society: The secret algorithms that control money and information*. Harvard University Press.

[111] Matteo Pasquinelli and Vladan Joler. 2020. The Nooscope Manifested Artificial Intelligence as Instrument of Knowledge Extractivism. *AI and Society*: 23.

[112] Jürgen Pfeffer, Katja Mayer, and Fred Morstatter. 2018. Tampering with Twitter's Sample API. *EPJ Data Science* 7, 1: 50.

[113] Jean-Christophe Plantin, Carl Lagoze, Paul N. Edwards, and Christian Sandvig. 2018. Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society* 20, 1: 293–310.

[114] Nikolaus Poechhacker and Severin Kacianka. 2021. Algorithmic Accountability in Context. Socio-Technical Perspectives on Structural Causal Models. *Frontiers in Big Data* 3. https://doi.org/10.3389/fdata.2020.519957

[115] Theodore M. Porter. 1996. *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press.

[116] Emilee Rader and Rebecca Gray. 2015. Understanding user beliefs about algorithmic curation in the Facebook news feed. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems,* 173–182.

[117] Inioluwa Deborah Raji and Joy Buolamwini. 2019. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (AIES '19), 429–435. https://doi.org/10.1145/3306618.3314244

[118] Kat Roemmich and Nazanin Andalibi. 2021. Data Subjects' Conceptualizations of and Attitudes toward Automatic Emotion Recognition-enabled Wellbeing Interventions on Social Media. *Proceedings of ACM in Human Computer* Interaction 5, CSCW2: 1-34.

[119] Legacy Russell. 2020. *Glitch Feminism: A Manifesto*. Verso.

[120] Jathan Sadowski and Roy Bendor. 2019. Selling smartness: Corporate narratives and the smart city as a sociotechnical imaginary. *Science, Technology, & Human Values* 44, 3: 540–563.

[121] Javier Sánchez-Monedero and Lina Dencik. 2020. The politics of deceptive borders: 'biomarkers of deceit' and the case of iBorderCtrl. *Information, Communication & Society* 0, 0: 1–18. https://doi.org/10.1080/1369118X.2020.1792530

[122] Christian Sandvig. 2015. Seeing the Sort: The Aesthetic and Industrial Defense of "The Algorithm" | NMC Media-N. *Media-N: Journal of the New Media Caucus* 10, 3. Retrieved August 14, 2020 from http://median.newmediacaucus.org/art-infrastructures-information/seeing-the-sort-the-aesthetic-and-industrial-defense-of-the-algorithm/

[123] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry*: 1–23.

[124] Michael Sawh. 2019. Getting all emotional: Wearables that are trying to monitor how we feel. *Wareable*. Retrieved August 28, 2020 from https://www.wareable.com/wearable-tech/wearables-that-track-emotion-7278

[125] Nete Schwennesen. 2019. Algorithmic assemblages of care: imaginaries, epistemologies and repair work. *Sociology of Health & Illness* 41, S1: 176–192. https://doi.org/10.1111/1467-9566.12900

[126] Irving Seidman. 2005. *Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences, 3rd Edition.* Teachers College Press, New York.

[127] Aaron Springer and Steve Whittaker. 2019. Progressive disclosure: empirically motivated approaches to designing effective transparency. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (IUI '19), 107–120. https://doi.org/10.1145/3301275.3302322

[128] Luke Stark. 2018. Algorithmic psychometrics and the scalable subject. *Social Studies of Science* 48, 2: 204–231. https://doi.org/10.1177/0306312718772094

[129] Luke Stark. 2019. Affect and Emotion in digitalSTS. *DigitalSTS: A Field Guide for Science & Technology Studies*: 117–135.

[130] Luke Stark and Jesse Hoey. 2020. The Ethics of Emotion in AI Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 782-793.

[131] H. Strömfelt, Y. Zhang, and B. W. Schuller. 2017. Emotion-augmented machine learning: Overview of an emerging domain. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction* (ACII), 305–312. https://doi.org/10.1109/ACII.2017.8273617

[132] Lucy Suchman. 2007. Human-machine reconfigurations: Plans and situated actions. Cambridge University Press.

[133] Latanya Sweeney. 2013. Discrimination in online ad delivery. *Queue* 11, 3: 10.

[134] Keeanga-Yamahtta Taylor. 2019. *Race for profit: How banks and the real estate industry undermined black homeownership.* UNC Press Books.

[135] Zeynep Tufekci. 2014. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *Eighth international AAAI Conference on Weblogs and Social Media.*

[136] Zeynep Tufekci. 2015. Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colo. Tech. LJ* 13: 203.

[137] Kristen Vaccaro, Dylan Huang, Motahhare Eslami, Christian Sandvig, Kevin Hamilton, and Karrie Karahalios. 2018. The Illusion of Control: Placebo Effects of Control Settings. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 1–13. https://doi.org/10.1145/3173574.3173590

[138] Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2020. "At the End of the Day Facebook Does What ItWants": How Users Experience Contesting Algorithmic Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2: 167:1-167:22. https://doi.org/10.1145/3415238

[139] Nicholas A. Valentino, Ted Brader, Eric W. Groenendyk, Krysha Gregorowicz, and Vincent L. Hutchings. 2011. Election Night's Alright for Fighting: The Role of Emotions in Political Participation. *The Journal of Politics* 73, 1: 156–170. https://doi.org/10.1017/S0022381610000939

[140] José Van Dijck. 2014. Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12, 2: 197–208.

[141] 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*, 63–76.

[142] Rick Wash. 2010. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security* (SOUPS '10), 1–16. https://doi.org/10.1145/1837110.1837125

[143] Christian von der Weth, Ashraf Abdul, Shaojing Fan, and Mohan Kankanhalli. 2020. Helping Users Tackle Algorithmic Threats on Social Media: A Multimedia Research Agenda. In *Proceedings of the 28th ACM International Conference on Multimedia* (MM '20), 4425–4434. https://doi.org/10.1145/3394171.3414692

[144] Ben Williamson. 2018. Silicon startup schools: technocracy, algorithmic imaginaries and venture philanthropy in corporate education reform. *Critical Studies in Education* 59, 2: 218–236.

[145] Langdon Winner. 1980. Do artifacts have politics? *Daedalus*: 121–136.

[146] Richmond Y. Wong, Deirdre K. Mulligan, and John Chuang. 2017. Using science fiction texts to surface user reflections on privacy. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 213–216.

[147] Allison Woodruff, Sarah E. Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. 2018. A Qualitative Exploration of Perceptions of Algorithmic Fairness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 1-14. https://doi.org/10.1145/3173574.3174230

[148] Meredith GF Worthen. 2014. An invitation to use craigslist ads to recruit respondents from stigmatized groups for qualitative interviews. *Qualitative Research* 14, 3: 371–383.

[149] Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R. Zaiane. 2017. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)* 50, 2: 1–33.

[150] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-12.

[151] Biqiao Zhang and Emily Mower Provost. 2019. Automatic recognition of self-reported and perceived emotions. In *Multimodal Behavior Analysis in the Wild*. Elsevier, 443–470.

[152] Shoshana Zuboff. 2015. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30, 1: 75–89.

[153] Shoshana Zuboff. 2019. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. Public Affairs, New York.

Received January 2021; revised July 2021; accepted November 2021.