Reflective Design for Informal Participatory Algorithm Auditing: A Case Study with Emotion AI

Noura Howell
Watson Hartsoe
Jacob Amin
Vyshnavi Namani
Georgia Institute of Technology
Atlanta, Georgia, US

ABSTRACT

This paper suggests how reflective design can aid informal participatory algorithm auditing. Drawing from reflective design, we designed a simple web-form probe to invite critical reflection on Emotion AI, ethically controversial techniques predicting individuals' emotions. Participants engaged the probe throughout their daily lives for about a week. Then, we interviewed participants about their experiences and reflections. Our findings surface themes around participants' (i) critiques of Emotion AI, (ii) factors contributing to inaccuracy, and (iii) patterns of miscategorization. Our discussion contributes (1) recommendations for Emotion AI and (2) how reflective design may offer considerations to inform algorithm auditing. Overall, our paper suggests ways critically-oriented design research can engage AI ethics through informal, participatory, exploratory algorithm auditing.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Emotion AI, reflective design, algorithm audit

ACM Reference Format:
Noura Howell, Watson Hartsoe, Jacob Amin, and Vyshnavi Namani. 2024. Reflective Design for Informal Participatory Algorithm Auditing: A Case Study with Emotion AI. In *Nordic Conference on Human-Computer Interaction (NordiCHI 2024), October 13–16, 2024, Uppsala, Sweden.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3679318.3685411

1 INTRODUCTION

This paper engages algorithm auditing and Emotion AI through design research methods: We used reflective design to investigate ethical issues of Emotion AI by inviting participants' interpretations of and critical reflections on Emotion AI. Through this, we identify the potential for reflective design to inform informal participatory algorithm auditing.

NordiCHI 2024, October 13–16, 2024, Uppsala, Sweden

ACM ISBN 979-8-4007-0966-1/24/10
https://doi.org/10.1145/3679318.3685411

Algorithm audits support identifying, addressing, and resisting ethical issues of AI harms, and they help reap potential benefits of AI systems. Algorithm audits help identify errors and discriminatory bias in algorithmic systems [97], which is essential for identifying vectors of harm and improving accuracy and fairness. HCI is currently exploring new approaches for more participatory algorithmic auditing, leveraging how everyday users can bring their unique lived experiences and identities to surface what is unknown to AI developers [41, 128].

One area where more algorithmic audits are urgently needed is Emotion AI. Emotion AI is a rapidly expanding set of technologies that predict individuals' emotions based on data; in this paper, we focus on Emotion AI based on analyzing images of facial expressions (Facial Expression Analysis or FEA [7]). Emotion AI can support well-being (e.g., [4, 80, 120]), but it is also deployed for harmful surveillance in education [42], hiring [117], work [28, 35, 118], and other areas [38, 93]. Some prior work audits Emotion AI (e.g., [50, 79, 115]), but errors with one algorithm may not necessarily be the same as errors with another algorithm, and further audits are needed in this rapidly expanding technology landscape.

This paper investigates ethics of Emotion AI by inviting participants to experience Emotion AI firsthand and critically reflect on it. As designers, drawing from reflective design approaches [124], we developed a technology probe [26] as an elicitation device, an interactive artifact designed to prompt reflection and help study participants' experiences and perspectives. Our probe is a simple mobile-friendly web form that asks participants to self-report their own emotions, then shows them how Emotion AI categorizes their emotions. We used an existing Emotion AI, Morphcast [3], chosen for its more protective privacy policy, and displayed Morphcast's predictions unaltered; we explain these design decisions in Sec. 3. Participants used the probe throughout their daily lives, prompting reflections situated in varied contexts. After about a week engaging the probe, we interviewed participants about their experiences and reflections.

Our findings detail participants' experiences with Emotion AI predictions: (Findings Sec. 5.1) Participants critically reflected on what they describe as fundamental issues of Emotion AI, (5.2) identified factors that they believed contributed to inaccuracy of Emotion AI predictions, and (5.3) described patterns of miscategorization in which Emotion AI predictions were repeatedly inaccurate.

Our discussion offers: (Discussion Sec. 6.1) recommendations for designing Emotion AI, and (6.2) considerations for how reflective design can inform informal participatory algorithm auditing.

Overall, this paper offers contributions for design research, algorithm auditing, and designing Emotion AI: For design researchers, our paper suggests opportunities for using existing design research approaches to support identifying algorithm errors as part of broader efforts in participatory algorithm auditing. For algorithm audit researchers, our paper offers considerations for how drawing from reflective design may be helpful for inviting participants' holistic, critical reflections during informal participatory algorithm audits. For designing Emotion AI, this paper offers design recommendations around the risk of stigmatizing labels, supporting more realistic knowledge claims, and when the implication is not to design. Overall, our paper builds bridges between approaches in AI ethics and design research, suggesting promising directions for both.

2 BACKGROUND

We connect Emotion AI, reflective design, and algorithm auditing.

2.1 Emotion AI

Emotion AI is controversial. Proponents expound benefits for well-being and security, while critics decry ethical risks, cultural bias, and scientific flaws. Emotion AI is increasingly deployed in high-stakes contexts such as hiring, education, and security [38, 93]. An urgent need to continue investigating Emotion AI ethics from multiple angles motivates this paper.

Emotion AI predicts human emotions and related psychological characteristics [133]. Affective computing [106] laid the foundation for Emotion AI. In this expansive space, closely related approaches are also called emotion recognition (e.g., [13]), emotional biosensing (e.g., [71]), or terms around physiological signals, biosignals, or biodata. Emotion AI can use various biodata; we focus on Emotion AI using facial expression analysis [86], popular due to widely available image data.

2.1.1 Benefits. HCI advances Emotion AI for well-being. A recent NordiCHI workshop explored possibilities for more empathic interfaces via affective computing [27]. Emotion AI and emotionally-pertinent biodata could enhance empathetic communication [39, 51, 92, 119, 120], support happiness, productivity, and reduce stress [70, 80, 98], or even help parents understand children's emotions during video-game-playing [105]. Displaying predicted emotions could support team satisfaction [134], learning [49], and more inclusive meetings [121]. Emotion AI consumer products target well-being and self-care [5, 8, 11, 12, 141, 145]. Sanches et al.'s HCI synthesis on affective health includes possibilities for emotional data to support wellbeing [123], as does Slovák et al.'s HCI synthesis on emotion regulation [131]—both calling for integration with broader care systems [123, 130]. We also highlight decolonial mental health [104] and ethical guidelines for Emotion AI [96].

Overall, Emotion AI's potential to support well-being seems to motivate continued HCI research on Emotion AI despite grave ethical risks.

2.1.2 Ethical risks. Emotion AI deployments raise ethical risks in surveillance [38, 93], education [42, 59, 62, 94], workplaces [28, 35, 67, 108, 118], preemptive threat detection [9, 59, 76], evaluating job candidates [6, 10, 32, 63, 110, 117], and security [22, 38]. Stark et al. outline 'toxic' social implications if people internalize algorithmic models of emotion [135]. They critique Emotion AI in education [42] and trace historical links to racist pseudoscience [136].

Overall, Emotion AI raises many ethical risks. Alongside many other factors, risks of harm depend on how people interpret Emotion AI predictions, and what actions people take based on those interpretations. This motivates our study's focus investigating how people interpret Emotion AI predictions. Our study mitigates ethical risks by showing people only Emotion AI predictions about themselves and not suggesting any actions.

2.1.3 Cultural bias. Emotion AI can embed cultural bias. Affective computing aims to detect categories of emotion, claiming to transcend cultural context [33]. Sengers, Boehner, and collaborators draw from cultural anthropology to critique these claims as culturally reductive. They offer design tactics for affective computing to embrace nuance, ambiguity, and diversity of emotion as assets enriching the design of computational systems [23–25, 125, 127]—drawing from reflective design by Sengers et al. [124].

Our past work builds on this lineage to advance approaches for design researchers to engage computational ways of knowing emotion more ethically [71–74].

2.1.4 Scientific flaws. Emotion AI's scientific basis is critiqued [64]. Emotion AI often uses Ekman's theory of universal, discrete emotions inferred from facial expressions [21, 46, 47, 78]. However, Feldman Barrett et al. identify foundational scientific weaknesses in this [17]. Feldman Barrett critiques overblown Emotion AI claims in popular press, and calls for further scientific inquiry into complex relationships between emotional experiences and facial expressions [34]. She explains, *"[Companies] can detect a scowl, but that's not the same thing as detecting anger"* [149].

So, we designed our study to give participants firsthand experience with the difference between facial expressions (e.g., 'detecting a scowl') and underlying emotions (e.g., 'detecting anger') (Sec. 4).

2.1.5 What do laypeople think of Emotion AI?. HCI studies data subjects' perceptions of Emotion AI. 'Data subjects' refers to those subject to data collection, drawing from Roemmich, Andalibi, et al. [116]. Andalibi and coauthors highlight data subjects' ethical concerns with Emotion AI on social media [116]; regarding accuracy, privacy, transparency, and contestability [60]; as threatening autonomy [13] and agency [60]; and in workplaces [35, 118]. They also critically analyzed Emotion AI hiring services [117] and Emotion AI patents [28]. A survey ($N > 3500$) found most rated Emotion AI as acceptable for advertising (framed as low-stakes) but unacceptable for hiring (high-stakes) [48]. McStay found people under age 34 rated emotion recognition as more acceptable than those over 34 [95].

Prior work describes how people conceptualize Emotion AI impacts, *without* participants directly experiencing Emotion AI. Motivated by this, our project invited participants to critically consider Emotion AI, grounded in firsthand experience with Emotion AI.

2.1.6 Summary. Emotion AI's potential to support wellbeing seems to motivate many researchers to continue developing Emotion AI despite ethical risks (Sec. 2.1.1). Thus, even though Emotion AI poses grave ethical risks (Sec. 2.1.2), embeds cultural bias (2.1.3), and has scientific flaws (2.1.4), it seems likely that Emotion AI will continue to be an area of ongoing research and development. This

motivates a need to continue investigating Emotion AI ethics and risks from many angles, and to more deeply integrate AI ethics into developing Emotion AI.

Responding to this need, recent workshops at DIS [137] and NordiCHI [143] bring together designers working with biodata (including Emotion AI) for discussions of AI ethics. Also responding to this need, our paper explores using reflective design to stitch connections between designing and auditing Emotion AI, for investigating Emotion AI ethics.

2.2 Reflective design

Reflective design encourages designers and participants to critically reflect and question the status quo and embedded assumptions in technology, society, participants, and designers (Sengers et al. [124]). Drawing from critical theory, they define reflection as "critical reflection, or bringing unconscious aspects of experience to conscious awareness, thereby making them available for conscious choice" [124, p. 50]. More generally, 'reflection' is often a design goal, albeit with varying conceptual precision [19]. Particularly relevant to our project, a key principle of reflective design is that "Technology should support skepticism about and reinterpretation of its own working" [124, p. 55]. We draw from reflective design in our project to prompt critical reflection on Emotion AI.

As reflective design examples, Affector presents an abstract, ambiguous affect display, probing open-ended interpretation and reflection [126]. Kaye's "I just clicked to say I love you" project developed a minimalist desktop dot as a Virtual Intimate Object (VIO) for long-distance couples, using a probe with short-answer text responses inviting participants to share reflections [81]. Reflective design has long been used to critically engage computational ways of knowing or displaying affect and emotion (e.g., Affector [126]) and used simple yet effective UI and probes (e.g., VIO [81]). In our project, we designed a technology probe [26] inviting participants to experience and critically reflect on Emotion AI (Sec. 3).

2.2.1 Critically-oriented design research draws from reflective design. Since reflective design's 2005 introduction, myriad critically-oriented, speculative, discursive, and related design approaches have flourished [15, 37, 43, 44, 77, 85, 107, 138]. Without collapsing the richness of different approaches here, many approaches draw from reflective design's move away from solutionism toward encouraging reflection, critique, debate, and reimagining alternatives. Many critically-oriented design research projects critique and reimagine computational ways of knowing emotion. For example, Tsaknaki et al. critically reimagine emotion, biodata, and technology [144] using Søndergaard et al.'s fabulation as an approach for design futuring [77]. Sanches et al. offer ways to apply the feminist new materialist concept of *agential realism* to design alternatives with biodata including emotional biodata [122]. *Reflective informatics* outlines dimensions of designing for reflection specifically for information: breakdown, inquiry, and transformation [18]–we draw from this in designing our probe (Sec. 3).

2.2.2 Reflective design critiques computational ways of knowing emotion. Prior work connects reflective design and Emotion AI: Sengers, Boehner, and collaborators introduced reflective design as a method [124]. Boehner, Sengers, and collaborators critique computational approaches to knowing emotion underlying Emotion AI, in part through reflective design [24, 124–127]. Reflective design as a method is well suited for prompting critical engagement with Emotion AI.

2.2.3 Summary. Our paper stitches further connections between reflective design and more recent ways of engaging Emotion AI ethics via auditing. This responds to a need (summarized in 2.1.6) to investigate Emotion AI from many angles, including by bringing design research approaches to bear on AI ethics. Recent work at NordiCHI shows the promise of applying critically-oriented design research methods for critical algorithmic engagement, such as Klumbytė et al.'s work connecting critical design with algorithmic experience [84]. Related but distinct, we choose to draw from reflective design because of reflective design's foundational call to make critical reflection a key part of technology design, and its lineage of already critiquing computational approaches to emotion.

2.3 Algorithm audits

Algorithm audits evaluate algorithms' behaviors, often detecting systematic errors or societal bias. For example, Buolamwini and Gebru's foundational Gender Shades audit revealed intersectional racist and sexist bias in facial recognition [31]. In response, companies improved their algorithms [113].

2.3.1 Audits of Emotion AI. We focus on Emotion AI that analyzes facial expressions (Facial Expression Analysis, FEA) (e.g., [1, 7, 150]). These algorithms are related to, but distinct from, facial recognition algorithms, and so different audits are required. For example, an Emotion AI algorithm could miscategorize a yawn as a gape of surprise [79] or flatten cultural nuance. Documenting Emotion AI errors may support algorithmic contestability, the ability to contest algorithmic knowledge claims [68, 146].

Emergent audits of Emotion AI suggest potential biases. Kaur et al.'s workplace study, comparing Emotion AI predictions to participant self-report, found alignment under 60% [79]. Some Emotion AIs identify joy and surprise better than anger, fear, and sadness [87]. Disturbingly, one audit found Black faces were categorized as angrier than white faces [115]. Another audit found some parts of facial expressions were better recognized for Black faces [101]. Images of females, African Americans, and people over 40 scored higher on smiling-related metrics [50], while another study found lower performance for people over 34 [101]. Emotion AIs tend to perform worse at detecting components of facial expressions around the eyes when people are wearing glasses [101].

Because facial recognition and Emotion AI algorithms are related but distinct, we can hypothesize that Emotion AI might inherit documented racist and sexist bias of facial recognition algorithms; further Emotion AI audits should evaluate this. Wearing sunglasses blocks visuals of the face, reducing Emotion AI accuracy [102]. We can speculate that Emotion AI could present additional issues around glasses, hats, hair, or face coverings—which could by proxy indicate health, religious, or regional characteristics—as well as cultural or ableist norms of how or when it is appropriate to perform different facial expressions. These are speculative hypotheses; further audits of Emotion AI are needed.

2.3.2 Limitations of auditing. Algorithm audits can be powerful tools for accountability (e.g., Radiya-Dixit and Neff [109]), yet audits have limitations. Raji et al. [112] explain audits are not a complete solution and do not guarantee future algorithm deployments will proactively consider identified issues. Audits can become outdated as new versions of algorithms are deployed [97]. Internal processes for algorithm audits can risk serious conflicts of interest [129, 152]. Audits evaluate how well an algorithm does what it claims to do, but they do not evaluate whether an algorithm *should* do that [82, 97]. With this in mind, we explore how reflective design might invite more holistic critiques, including an opportunity to consider what an algorithm should or should not do.

Audits are often done by experts [97]. Yet, as Deng et al. highlight, groups marginalized from technical expertise tend to be both disproportionately impacted by algorithmic bias *and* uniquely well positioned to identify algorithmic bias [40]. Costanza-Chock, Raji, and Buolamwini call for directly involving those harmed by AI in auditing [36]. Vecchione, Barocas, and Levy call for community-engaged audits [147]. This motivates work on participatory algorithm auditing.

2.3.3 Participatory algorithm auditing. User-driven audits have emerged to further uncover algorithmic discrimination. Users have independently identified racist and sexist bias in deployed algorithms from Google, Twitter, and Apple [61, 66, 148]. Shen et al. describe this kind of *everyday algorithm auditing* as "a process in which users detect, interrogate, and understand problematic machine behaviors via their daily interactions with algorithmic systems" [128, p. 2]. Li et al. analyzed user-driven auditing on Twitter and recommend designing for discussion, deliberation, and greater access/visibility of algorithms [91]. DeVos et al. studied how participants identify bias in Google image search, finding participants' prior experiences with bias were key influences in how they searched for and identified algorithmic bias [41]. Deng et al. explore how AI industry practitioners can engage end-users in auditing [40], suggesting deriving actionable insights, soliciting critical holistic feedback, scaffolding users' auditing, and recognizing the importance of auditors' identities. We return to these considerations for participatory algorithm auditing in the discussion.

Some systems aim to support user-driven auditing. Beat The Machine invites users to find examples that a model will misclassify [14]. Dynabench invites users to provide text examples to 'fool' an NLP model [83]. Search Atlas shows Google search results as they appear to users in different countries, inviting comparison and questioning [100]. IndieLabel helps users create audit reports and surface issues, used in content moderation toxicity modeling [89]. Our discussion unpacks ways that our probe supported user-driven auditing, especially engaging calls from Deng et al. on engaging end-users in auditing [40].

3 DESIGNING THE PROBE

As our original motivation as design researchers, we designed a probe [26] to invite participants to experience and critically reflect on Emotion AI in their daily lives. Prior work studies how people conceptualize present-day Emotion AI benefits and harms based on more general descriptions of Emotion AI writ large (Sec. 2.1.5). Distinct from and complementing this prior work, we designed our probe to invite participants' critical reflections on Emotion AI grounded in their firsthand experience with a particular Emotion AI system. Since cultural probes were introduced [54, 57], design research has adapted probes in many ways (e.g., [75, 88, 103, 142]). We draw from Boehner et al.'s synthesis of probes in HCI [26].

Our probe asks participants to compare self-report of their emotions with an Emotion AI system's categorization/prediction/label(s) of their emotions in the same moment. It consists of a simple mobile-friendly web form:

(1) Page 1 asks the participant to self-report their feelings and context.
   (a) What is your first name?
   (b) Take a photo of yourself that includes a front view of your visible face.
   (c) Take a photo of your surroundings.
   (d) In a few words, describe what is happening around you.
   (e) In a few words, how are you feeling?
(2) Page 2 shows how an Emotion AI system categorizes their emotions, and asks participants to react to this.
   (a) In a few words, describe how the Emotion AI is classifying your emotion.
   (b) How would you rate the accuracy of how the Emotion AI analyzes your emotion? (on a scale of 1 to 5)
   (c) Overall, what is your immediate reaction or impression in response to seeing how the Emotion AI analyzed you?

The first page asks participants to log a photo of their face, a photo of their surroundings, a few words describing their location or activity at that moment, and a few words describing their emotions at that moment. The second page shows participants how the Emotion AI categorizes their emotions in that moment, and asks participants to rate the accuracy of the Emotion AI's predictions on a scale of 1 to 5, and share their immediate reaction or impression in response to how the Emotion AI analyzed them. We do not include screenshots because the visual design of the web form was not a focus of our design effort, nor was it of particular interest to participants. Rather, in designing the probe, we focused on structuring this particular 'information sequence' of gathering information from participants and presenting Emotion AI predictions back to participants, to offer firsthand experience with Emotion AI and prompt critical reflection Emotion AI.

3.1 Drawing from reflective design

Here, we explain how the design of our probe draws from reflective design by linking particular principles and tactics of reflective design [124], as well as considerations for reflective informatics [18], to particular design decisions.

The probe is designed with the reflective design principle that "Technology should support skepticism about and reinterpretation of its own working" [124, p. 55]. We designed the probe to invite skepticism about Emotion AI, asking participants to rate the accuracy of the Emotion AI's predictions (the second question on page 2 of the web form, as outlined above). This invites participants to consider whether they view the Emotion AI prediction as accurate or inaccurate, thus inviting skepticism. The probe also designs for reflection "as an integral part of experience" [124, p. 56] by requesting participants' to share their reactions and reflections each time

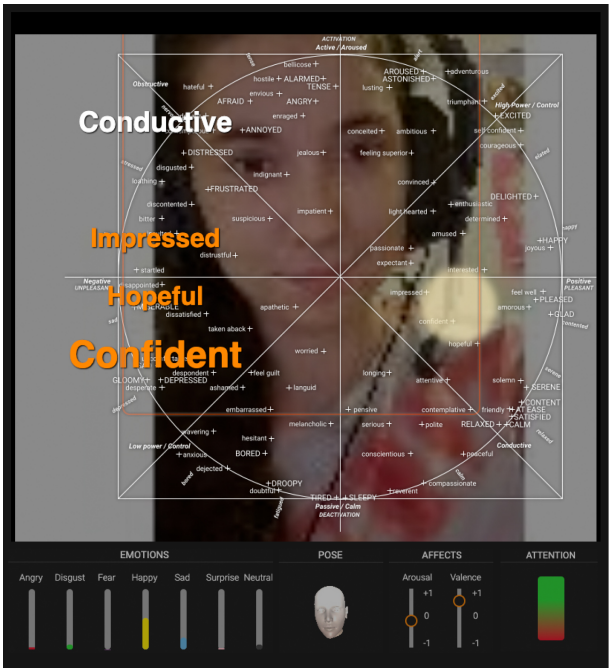

Figure 1: Participants saw how Morphcast [3], the Emotion AI system used by our probe, categorized their emotions. Our probe embedded a view of Morphcast's display, unaltered, not designed by us. We did this to prompt reflection [124] by leveraging tactics of ambiguity [56]: "over-interpret data" [56, p. 238] and "expose inconsistencies" [56, p. 238]. Morphcast's emotional adjectives (orange) often 'over-interpret' emotions, seeming 'inconsistent' with participants' self-report, inviting critical reflection. How our probe draws from reflective design is detailed in Sec. 3.1. Again, we did *not* redesign Morphcast's display; we embedded Morphcast's display in the probe. Although our prior work has designed displays of emotional biodata [72–74], in this project we presented Morphcast's display unaltered because our aim is to study how participants experience present-day Emotion AI as it already exists in commercially available products such as Morphcast.

they use the probe (the third question on page 2 of the web form, as outlined above). One strategy of reflective design is to "Build technology as a probe" [124, p. 56] to study people's experiences and reflect on technology design itself.

Another strategy of reflective design is to "Invert metaphors and cross boundaries" [124, p. 57]. The information sequence we designed critically inverts the typical information flow of Emotion AI systems. Typically, Emotion AI systems seek to use Emotion AI's emotion predictions to supplant or lessen the need for participant's emotion self-report; the aim is for Emotion AI predictions to be useful as the primary source of information and the self-report emotions to become secondary. By asking participants to describe their own emotions before viewing Emotion AI's predictions, we position self-reported emotions as primary and Emotion AI predictions as secondary.

Drawing from reflective informatics [18], which builds on reflective design specifically for informatics, the probe invites moments of *breakdown* and *inquiry*, dimensions of reflective informatics. Based on our own informal use experiences with the Emotion AI, we anticipated that participant emotion self-report would often differ significantly from the Emotion AI predicted labels—these errors are moments of *breakdown* of the Emotion AI. The probe invites participants to *inquire* about the similarities or differences in how they describe their own emotions and how the Emotion AI categorizes their emotions. Intentionally exposing participants to Emotion AI errors also responds to Raji et al.'s point that too often AI ethics debates assume AI systems are functional [114].

We designed the probe to require participants to take a photo of themselves so that the Emotion AI can analyze it. Although many Emotion AI systems claim that by unobtrusively monitoring—or, in less flattering terms, surveilling—people, they are able to 'capture' more 'honest' facial expressions and inferred emotions, we intentionally ask participants to consciously choose to submit a photo of themselves to Emotion AI. We do this for ethical reasons, so that our system does not replicate or normalize patterns of Emotion AI surveillance, and instead promotes an ongoing proactive consent and awareness each and every time the participant submits a photo for Emotion AI analysis. This also prompted people to reflect on the presence of Emotion AI at various moments and contexts throughout their daily lives.

We designed the probe to ask participants to provide a photo of their surroundings, so that later in the post-interview we could show this photo back to participants to help them remember the surrounding context in which the log entry took place. We supplemented this by requesting a few words describing their activity or location, again to help recall the context of the log entry.

After the probe shows the participant an Emotion AI prediction based on their photo, it asks participants to log their subjective rating of the Emotion AI's accuracy for that prediction. We designed the probe this way to continually prompt participants to critically evaluate the in/accuracy of Emotion AI throughout many varied contexts of their daily lives. We also added a catch-all text-entry for any additional thoughts.

We designed the probe to be convenient to use on a mobile device, robust to consistently function across many different mobile devices owned by participants, and secure in handling participant data. The information provided by the participant is handled as an embedded Qualtrics form. Qualtrics hosts surveys, allowed by our IRB as a secure way of handling human subjects data. It uses the device's camera and the Morphcast API for Emotion AI [3]. It displays a subset of six emotion categories determined by Morphcast, representing the probability that the participant's facial features corresponded to the shape of disgust, anger, fear, happiness, sadness, surprise, or neutral, as well as additional less standard emotion descriptors provided by Morphcast, such as 'suspicious' or 'longing'. We chose Morphcast as the Emotion AI service because of their GDPR-compliant data policies and relatively affordable API.

3.2 Algorithm auditing emerged later

We did not intentionally design for algorithm auditing. Rather, this came through design emergence [55]. In embracing emergence as

Gaver et al. recommend, we adopt the strategy to "Present design research as a journey, not a quest" in which we as design researchers detail the story of the 'journey and return' of our process, and resultant outcomes or learnings [55, Section 3.3]. In presenting this 'journey', we set out by drawing from reflective design to foster critical reflection on Emotion AI with the probe—presented in this section. Our findings and discussion unpack the learnings upon 'return' around how the probe emerged as a way of supporting informal participatory algorithmic auditing.

4 STUDY

As design researchers, the original motivation of our study was to investigate participants' experiences, interpretations, and reflections on Emotion AI. We invited 22 participants to experience the probe, then interviewed each participant about their experiences and interpretations. (1) In an introductory 1:1 meeting, we introduced the study's aims, answered any questions, and obtained informed consent. We overviewed how Emotion AI works based on facial expression analysis, existing Emotion AI applications, and benefits/risks. We introduced the probe and requested participants log about 10 entries over a week. (2) Participants used the probe throughout their daily lives for about one week, varying from a few to over 15 entries. This aimed to ground participants in real-world capabilities of an existing Emotion AI system, with its persistent inaccuracies and biases. (3) During a second meeting, we interviewed each participant for about an hour about their experiences and interpretations with the probe and Emotion AI. We showed participants previous entries they had made with the probe, prompting them to recall and reflect upon how they were feeling in that moment vs. how Emotion AI categorized their emotions in that moment.

Our study differs from a typical human-centered or user feedback study of an artifact: We do not aim to improve the user experience with Emotion AI. We do not aim to improve the usability or design of the probe. Rather, we specifically wanted to understand participants' reactions to and interpretations of how Emotion AI predictively categorized their emotions, and we designed the probe and the study to focus on this. Our approach aligns with prior research studying participants' experiences with a technology probe [26] not to improve the design of the probe, but rather to investigate an area of interest.

4.0.1 Recruitment and demographics. We conducted the study (IRB-approved) online via Zoom to reach a broader range of participants. Participants had to be 18+, with a probe-compatible mobile device. We recruited via social media, newsletters, and flyers. In all, 22 people participated, located across the US. 11 use he/him, 12 use she/her, and 1 uses they/them pronouns. 12 were age 18-24, 6 age 25-29, and 4 age 30-39. They self-identified as Asian (4), South Asian (4), Chinese (1), South-East Asian/Indian (2), Black/African-American/Nigerian (1), Filipino/Spanish (1), Hispanic Latinx (1), White/Latino (1), Asian/White (4), White (2), or no response (1). During the post-interviews, participants often reflected on their own backgrounds and identity characteristics, such as being a person of color, a woman, on the Autism Spectrum, or having depression or PTSD. Additional participant demographics are in Appendix A, including any prior experience with AI.

4.0.2 Analysis. We used qualitative coding analysis. During the post-interview, participants' probe entries were revisited to prompt reflections. Interviews were transcribed and analyzed using qualitative coding using abductive analysis [139]. We brought in no predefined codes, letting codes emerge from the data, and did two rounds of coding. All four authors took part in a first pass of open coding, where each interview was analyzed by at least two researchers. We discussed overlaps and differences in our codes and interpretations to reach consensus and uncover insights. We clustered codes into emergent themes, also noting when participants' experiences notably differed from a cluster. Then, in a second round of coding, we focused on instances when participants critiqued Emotion AI or described moments when they felt Emotion AI incorrectly predicted their emotions.

Abductive analysis combines grounded theory with researchers' own social and intellectual perspectives to surface surprising findings [139]. Abductive analysis calls for greater researcher reflexivity on how researchers' positionality and intellectual influences shape their analysis, and shifts away from summarizing patterns to highlighting differing perspectives that may be surprising [139]. In this sense, rigor comes not through mitigating researchers' cognitive bias, but rather through acknowledging and reflecting on how researchers' positionality and perspectives necessarily influence the analysis. This resonates with diffractive analysis [52, 53, 58], taken up in design research for qualitatively analyzing participant experiences (e.g., [111]). Diffractive analysis places similar emphasis on researcher reflexivity and highlighting different perspectives. Our intellectual influences are summarized in Background (Sec. 2). Our positionalities motivate attention to critiquing Emotion AI bias, and we highlight sometimes unusual or surprising perspectives from a minority of participants.

4.0.3 Positionality. Currently at a US university, we come from different cultural, racial, and religious backgrounds. As immigrants or children of immigrants from outside the global North/West, we were attuned to cultural differences in emotional expression and Emotion AI's potential to embed cultural bias. The lead author grew up in a multicultural family scattered across North America and the Middle East—experiencing firsthand how emotional expression varies culturally. She also witnessed xenophobia, sensitizing her to how technology can reinforce stigma, stereotypes, or otherization. This shapes our interest in challenging universalizing claims of Emotion AI.

5 FINDINGS

Participants critiqued Emotion AI (Sec. 5.1), suggested factors seemingly contributing to inaccuracy (Sec. 5.2), and described patterns of miscategorization (Sec. 5.3). Throughout, we use 'categorize', 'predict', and 'label' as synonyms for Emotion AI's outputs.

5.1 Critiquing fundamental issues with Emotion AI

Participants fundamentally critiqued how Emotion AI conceptualized or predicted emotion.

5.1.1 Facial expressions are not the same as felt emotion. Some participants articulated a distinction between facial expressions

"I am playing Minecraft at my desk while chatting with friends" ... feeling "Content and accomplished" (P2)

"I am working while lounging on the couch" ... feeling "Better than an hour ago" (P5)

"Living room, watching TV" ... feeling "Calm and happy" (P6)

"Done with half part of assignment" ... feeling "Looking forward z excited" (P13)

"Talking to my friend" ... feeling "Happy, Excited" (P3)

Figure 2: With our probe, participants saw how Emotion AI algorithm Morphcast [3] categorized their emotions (above images), juxtaposed with their emotion self-report (text below). We intentionally showed participants' Morphcast's Emotion AI display unaltered to prompt critical reflection. These design decisions are elaborated in Sec. 3 and Fig. 1. Our protocol was IRB-approved and participants consented to publishing photos.

and felt emotions. P9 said, *"Obviously, physical presentation does not always equal how you're feeling."* So, one's facial expression, a form of physical self-presentation, cannot be equated with one's inner feelings. P4 said, *"People don't really express emotions on their face. If I feel happy, it doesn't really show. Unless it's instantaneous, like a reaction—then the emotion will show—but if it's, like, constant emotion, it doesn't really show."* In other words, P4's immediate reactions might show as facial expressions, but a more constant or continuing emotion would not. P4 also pointed out that *"People can fake emotions,"* such as via facial expressions. Participants' recognition of this distinction reinforces critiques of Emotion AI from prior work (Sec. 2.1.4).

5.1.2 Emotions can be unrelated to one's present-moment context. One participant noted a challenge of interpreting Emotion AI predictions: a person's emotions may be unrelated to their immediate surroundings. Regarding a hypothetical Emotion AI system flagging negative emotions in students at school, P1 suggested, *"asking [students] if anything's happening outside of school, [because] it's not about the lesson."* In other words, students' emotions in school could be unrelated to their immediate classroom context and instead related to broader issues. So, even if Emotion AI correctly predicted people's emotions in a particular context, it is invalid to assume those emotions are related to that context. This is an important distinction to acknowledge because, too often, Emotion AI applications assume an emotion is directly related to the surrounding context of interest to the application (e.g., [38, 93]).

5.1.3 Some emotions may be too complex to be detected by facial analysis. Some participants critiqued some Emotion AI predictions as too complex to be feasible. When it predicted feeling 'guilt', P21 critiqued, *"Feeling guilt sort of feels like a very complex emotion to assess based on someone's expression."* P21 considers it a stretch to predict a complex emotion such as guilt based on facial expression. P13 questioned the scientific validity of some predictions, commenting *"These [predictions] are things that I'd probably understand about another person after actually looking at their behavior, looking at the way they conduct themselves over a duration of time. How can I just look at a face and... like, classifying... I have a tough time understanding this as scientific."* P13 suggests some predictions made by Emotion AI are impossible to infer from facial expressions.

5.1.4 Summary. Participants' critiques point to fundamental issues with Emotion AI, issues already discussed in ongoing research. In this sense, the design of our probe and study facilitated participants' critical reflection on Emotion AI. Participants' critiques matter because, too often, Emotion AI applications make over-reaching claims about what can be inferred from facial expression [17]; in response, design researchers have called for greater humility [71] and contestability [68] with Emotion AI knowledge claims.

5.2 Factors contributing to inaccuracy

Participants described a variety of factors that seemed to contribute to inaccurate Emotion AI predictions. As Emotion AI is increasingly proposed and deployed for everyday contexts, it is essential to consider factors influencing accuracy.

5.2.1 Lighting. Participants observed that predictions seemed less accurate in dimmer or darker lighting. P21 observed, *"the system did quite poorly in dim lighting,"* and then he *"started trying to take photos more in bright light."* Other participants made similar observations. As P10 put it, in darker settings, *"it didn't capture as well."* Because the Emotion AI is based on analyzing photos of facial expressions, it relies on light for the image data. This aligns with known issues with facial analysis in dark settings: For example, Watkins' recounts a gig driver struggling with a facial recognition algorithm in a dark parking garage [151]. More specific to our case, Patel et al.'s literature review of Emotion AI based on facial imagery identifies illumination as a key factor impacting accuracy [102]. Here, our participants' reflections identify a known issue with Emotion AI based on facial imagery.

5.2.2 Glasses. Another factor participants expressed might influence accuracy was wearing glasses. P16 said, *"I think it might have been more accurate without the glasses. Because with the glasses, it*

was hard to change expressions, like it solidified to one thing very quickly. Without the glasses, it was more dynamic." P16 describes how when he was wearing glasses, the Emotion AI would quickly settle on one prediction, whereas when he was not wearing glasses, the predictions were more dynamic and seemingly more accurate. P11 observed, *"When I wasn't wearing my glasses it tended to be more accurate,"* reasoning that the glasses could have been *"an obstruction to my face,"* blocking visual perception of facial expression. On the contrary, P18 found wearing sunglasses did not seem to affect the Emotion AI, though they originally thought it might. Wearing glasses is known to impact accuracy of Emotion AI based on facial imagery [102]. Here again, our participants' reflections identify a known issue with Emotion AI.

5.2.3 Makeup. One participant reasoned eyeliner makes one's eyes look bigger, which Emotion AI might interpret as more awake. P2 described, *"Eyeliner makes your eyes look bigger, and bigger eyes generally means more awake."* P2 speculated that eyeliner could increase predicted alertness. Participants' reflections surface a suggestion for an algorithm audit investigating the impact of various makeup techniques on Emotion AI accuracy.

5.2.4 Social bias. Sometimes participants suggested potential links between Emotion AI's inaccuracies and axes of social discrimination. P5 said, *"My guess is that it just doesn't have data on people like me. So it's performing even worse than random on Indian people, people with a beard? That's my guess."* P5 has experience working with AI systems and considered that the Emotion AI may lack diverse training data, suggesting this could contribute to inaccuracy. P17 speculated that with Emotion AI there is *"probably racial bias, because not every race or culture expresses emotions the same way,"* suggesting racist and cultural bias in Emotion AI; P17 reported no prior experience with AI. Racist and sexist bias of facial recognition has become widely know [31]; it is reasonable to expect that Emotion AI reliant on facial imagery could suffer from similar biases. Indeed, audits of Emotion AI algorithms are beginning to uncover some evidence of racist bias (Sec. 2.3). Furthermore, emotional expression is deeply cultural and performative, with a diversity of ways of expressing emotions [25].

Participants noted potential ableist bias. P10, who self-identified on the Autism Spectrum, commented, *"There's not a lot of autistic people like me who are developing these algorithms,"* highlighting potential ableist bias in Emotion AI. This participant's reflection points to a need for Emotion AI to continue evaluating its suitability for neurodiverse populations (e.g., [90]).

Participants noted potential gender bias. P2 speculated, *"I would love to see what gendered reactions to this look like, because one of the things I've been thinking about a lot in the last couple years is gendered expectations of what it means to be camera-ready... I feel like there's a lot more pressure on women and female-presenting people to be camera-ready."* This participant's reflection Emotion AI's reliance on facial imagery could potentially embed sexist norms of visual self-presentation, again suggesting an avenue of further research.

Participants suggested potential ageist bias. P6 suggested, *"I think [Emotion AI] needs a lot of improvement. And I think it needs to do a lot more research with various age, various race people,"* pointing to the need for additional Emotion AI algorithm audits.

5.2.5 Summary. Participants' reflections surfaced issues that are relevant for Emotion AI algorithm audits. Participants' reflections surfaced factors contributing to inaccuracy and potential social biases with Emotion AI. In many cases, participants' reflections aligned with known issues impacting Emotion AI accuracy, issues already uncovered by Emotion AI audits of other algorithms. In some cases, participants' reflections pointed to areas that future audits could investigate more systematically. Overall, this shows how participants' reflections surfaced considerations highly relevant for Emotion AI auditing, underscoring the potential of reflective design to support algorithm auditing.

5.3 Patterns of miscategorization and attendant risks

Several participants observed patterns in which the Emotion AI repeatedly miscategorized their emotions.

5.3.1 Neutral baseline. Sometimes, participants connected this repeated miscategorization to thinking that their 'default' facial expression appeared more neutral. P20 said, *"I've been told I have a pretty neutral face... That was what I was getting all the time. Neutral. Regardless of what I was, how I was feeling."* P20 connected other people perceiving his face as 'neutral' to how the Emotion AI repeatedly miscategorized him as neutral regardless of his felt emotion. Kaur et al. [79] found their Emotion AI overwhelming classified all participants as neutral, in the workplace context of their study. Distinct from Kaur et al., participants used our probe throughout many varied contexts of their everyday lives beyond the workplace—although the probe itself may have been a somewhat 'neutral' interaction on their mobile device, other participants (in later paragraphs) experienced patterns of miscategorization beyond 'neutral'. Our findings show some participants experienced a pattern of being miscategorized as neutral, contrasting their self-reported emotions.

5.3.2 "Resting sad face" and imposing norms of facial expression. P2 shared that her neutral face often seems upset to other people, and reasoned the Emotion AI may have learned similar patterns and thus miscategorized her neutral face as sad. P2 described, *"Some people have different kinds of resting faces and people will make assumptions about that. I have learned that I have, like, resting upset face. If I make a neutral face or if I'm lost in thought, people will ask, 'Are you okay?' ... I think that's where a lot of that inaccuracy comes from, is that sense of, my neutral face apparently looks very upset to other people and probably matches patterns that this system has seen."* P2 shared anecdotes from her daily life in which she felt her face was neutral, but friends or teammates worried she was upset, which she attributed to "resting sad face strikes again". For P2, her "resting upset face" was something she already knew about herself, which she used to explain this miscategorization pattern.

For P11, it seems participating in the study prompted reflection on what her face might portray. She reflected, *"My face tends to portray, like, neutral and sad, which I didn't realize before this, but it seems like according to the AI I normally look sad. Even though I might not be feeling sad, but oftentimes the AI will tell me that I'm looking sad. It would get it correct if I was feeling sad... Sometimes I'd smile, and then it would tell me that I'm feeling happy. But then I*

would stop smiling, and then it would go back to saying sad. Because now that I think about it, I feel like it's not really an error with the AI. I think it's truly just what I look like, like my resting face, my baseline face, is more sad. And that's just kind of how it is. And because that's the case, AI isn't necessarily wrong. It's just maybe not, it just doesn't know the context of the situation. So it also can't tell it's not right." P11 observed that the Emotion AI frequently miscategorized her as feeling sad. Yet, P11 considered this not an error, reasoning that her 'resting face' appears sad, suggesting the Emotion AI was recognizing the appearance of sadness, even when that did not match her felt emotion. P7 similarly said that the Emotion AI frequently labelled them as sad, even when she was not feeling sad, and described how this caused her to *"doubt myself a little bit"* and worry she had particular facial features that the Emotion AI categorized as sad, and consider *"normative body standards"* which *"doesn't feel very positive"*. Whereas P2 already thought her resting face was sad, for P11 and P7 this was a new, somewhat unpleasant reflection prompted by Emotion AI's miscategorization.

Repeatedly miscategorizing participants as sad illustrates how Emotion AI risks imposing norms of facial expression and shaping self-perception in negative ways. While P2 described already knowing about her "resting upset face," P11 came to "realize" her neutral facial expression often appears sad. Similarly, P7 described often being miscategorized as sad, and said this made her wonder about *"normative body standards"* which *"doesn't feel very positive"*. Participants often linked patterns of miscategorization to suggesting how their face did not 'fit' with 'normative' facial expressions. These findings suggest Emotion AI could impose harmful algorithmic norms of facial expression.

5.3.3 'Disgust' as implying 'disgusting'. Some participants described how, when the Emotion AI predicted 'disgust' as their emotion, this was offensive. P5 described, *"It keeps telling me that there's a look of disgust on my face. And I find it offensive, to be honest, because I'm not actually sure if it's saying, 'oh, it's a look of disgust' or it's kind of starting to say, 'oh, no, you are disgusting', you know, because I'm not disgusted... Like, I'm not disgusted by bad software and, yeah, AI does have a tendency to be, you know, racially biased. So I'm starting to, like, now actually get upset."* The Emotion AI repeatedly miscategorized his emotion as 'disgust' when he was not feeling disgust*ed*, and P5 interprets this as implying he is disgust*ing*, finding this offensive and linking to potential racist bias.

Similarly, P18 observed the Emotion AI would often mislabel his emotions as disgust: *"Whenever I was not in a good mood, not having a good time, whatever it was, I always got a 'disgust' reading. And I'm like, 'I'm not disgusted. I'm just, like, not all that happy.' So, I don't know if I have, like, you know RBF [resting bitch face]... I don't know. It seems like sometimes my neutral, regular, slightly sad face was disgusted. And I don't know what was up with that."* In other words, the Emotion AI would miscategorize P18's neutral and negative expressions as disgust. P18 later added, *"I remember being particularly weirded out by it, just because in this moment, I wasn't in a good mood, but I wasn't feeling particularly much of anything. And it put me pretty strongly in the disgust category, which was a common theme, you know, as we've discussed previously. But this was one where I had a really strong reaction of, kind of [reacting] to the software, kind of, 'what the hell,' I'm like, 'this is just my face'."* P18 describes feeling how the Emotion AI miscategorized a somewhat neutral emotion as strongly in the disgust category, and it prompted a strong reaction. Similar to P5, P18 is offended at this miscategorization. Overall, miscategorizing as 'disgust' was a particularly offensive error. This points to the importance of identifying Emotion AI errors and assessing their impacts.

5.3.4 'Suspicious' as implying 'suspect'. When reviewing photos for which the Emotion AI predicted 'suspicious' and 'distrustful' by the Emotion AI, P16 remarked, *"I just see suspicious and distrustful as like my normal face;"* i.e., his normal face was being miscategorized as 'suspicious'. If such predictions were to be shared with police, he continued, *"that'd be scary. Because, like, if that's my normal neutral. And I know like, in general, I've gotten in trouble at school without doing anything, just because this teacher sees me, but I just have like guilty aspects."* P16 described how, in his daily life, he has been unfairly perceived as suspicious, and connects this to Emotion AI's miscategorization of him as suspicious. This points to harmful consequences if Emotion AI is used for assessing criminal risk.

6 DISCUSSION

Reflecting on our findings, we synthesize insights for designing Emotion AI (Sec. 6.1). Then, we offer considerations for how reflective design can inform informal participatory algorithm auditing (6.2). This helps stitch greater connections between critically-oriented design research and algorithm auditing, suggesting the beginnings of additional inroads for design research to impact AI ethics.

6.1 Insights for Emotion AI

6.1.1 Risk of stigmatizing labels. The patterns of miscategorizing people as feeling 'disgust' or 'suspicion', in addition to being inaccurate, implied offensive labels about the people themselves, as interpreted by some participants (Findings Sec. 5.3.3, 5.3.4). For example, some participants interpreted Emotion AI's prediction of their facial expression of 'disgust' as implying that they themselves were disgusting, or that labelling their facial expression as 'suspicious' and 'distrustful' implied that they themselves were suspected of wrongdoing and considered untrustworthy. Participants' interpretations of these miscategorizations as stigmatizing labels raise ethical concerns.

Stigmatizing labels (e.g., 'disgust', 'suspicious') could be especially harmful across power differentials and to historically marginalized groups; e.g., presenting Emotion AI predictions of a student to their teacher, a worker to their manager, or a job applicant to HR [38, 93]. Emotion AI and related biometric techniques are already used in criminal justice and security applications (e.g., [38, p. 50]). Algorithmic profiling of criminal/terrorist suspects can reinforce anti-Black, anti-Arab, and anti-Muslim oppression [29, 45]. If others interpret these Emotion AI labels similarly to these participants, then stigma—and perhaps even wrongful accusations—can harm people, spurred by Emotion AI miscategorizations. Beyond developers' intentions, how people interpret Emotion AI shapes its use in practice. **This underscores the importance of understanding how people interpret Emotion AI predictions.**

Our findings underscore that Emotion AI must design to mitigate its potentially stigmatizing effects. For designers and developers

of Emotion AI algorithms, we suggest that labels such as 'disgust' or 'suspicious' are particularly sensitive. These labels could be removed, renamed, required to meet a higher confidence threshold, or display additional guidelines for ethical interpretation. Overall, we call on Emotion AI systems—including algorithm and display—to more intentionally support users in carefully, ethically interpreting Emotion AI predictions.

6.1.2 More humble, realistic, contestable knowledge claims. Participants critiqued fundamental issues with Emotion AI (Sec. 5.1). Through interacting with the probe, they observed how facial expressions are not the same as felt expression (Sec. 5.1.1). This distinction echoes psychologist of affect Feldman Barrett's point that "[Companies] can detect a scowl, but that's not the same thing as detecting anger" [149], design research on emotion as socially performative [24, 25], and Kaur et al.'s participant saying, "I didn't know I looked angry" [79]. The distinction is important because, too often, Emotion AI applications conflate outer facial expressions and inner felt emotions. Our findings illustrate how many participants critiqued Emotion AI for implicitly conflating facial expression and felt emotion. **Rather than trying to 'overcome' the difference between facial expression and felt emotion, our findings underscore calls to recognize this distinction and design for it as inevitable.**

Participants also critiqued that one's emotions may not be directly related to one's immediate surrounding context (Findings 5.1.2). Acknowledging this distinction is important. Too often, Emotion AI applications assume an emotion is directly related to the surrounding context of interest to the application, such as assuming that someone's negative emotions indicate under-performance as a student or worker, or terrorist risk [2, 38, 93]. These inferences ignore the obvious possibility that someone may be unhappy about some aspect of their broader lives during school, work, or travel. This creates potential for harmful 'false positives' of wrongfully flagging someone as under-performing or a security threat. Emotion AI applications seem to embed this reductive logic to offer 'clear', 'actionable' insights [38, 93], but it leads to illogical claims, ethical harms, and—as our study found—critiques from users who identified this logical fallacy. Our work suggests that Emotion AI systems should design with this consideration in mind. Doing so could support not only more ethical, cautious interpretations of Emotion AI predictions, but also a better user experience.

Past work calls for designing for *contestability* with Emotion AI [68, 69, 146], and for making more humility in knowledge claims with computational ways of modeling emotion [71]. Our work underscores these calls and suggests how, contrary to typical values of designing tools to offer clear, simple, actionable insights, we join calls for designing Emotion AI as contestable, and for making more nuanced, humble knowledge claims with Emotion AI.

6.1.3 When the implication is not to design, but people build it anyway. Prior work highlights that, sometimes, the implication is not to design technology [20]. Some work calls for banning Emotion AI due to the potential for civil liberties violations (e.g., [93]). EU policy is shifting to ban Emotion AI in many contexts [65, 99]. Yet, it seems Emotion AI is not going away anytime soon. Emotion AI will likely continue to be used for surveillance for national security even in the EU [65], risking harm for criminalized minorities. On the other hand, Emotion AI could also have beneficial applications supporting mental health, if it designed with contestability, integrated with care systems, and proper protections as highly sensitive data [68, 69]. Until Emotion AI has vanished from research and practice, ethical engagements from multiple angles are needed—whether calling for bans [93], outlining ethical guidelines [96], or offering design insights for more thoughtful interpretation of Emotion AI.

6.2 Reflective design for participatory informal algorithm auditing

Leveraging reflective design approaches for our probe (Sec. 3.1) and study (Sec. 4) helped prompt participants to critically reflect on the algorithm, which resulted in participants' identifying algorithm errors. The probe asked participants to compare their emotion self-report with Emotion AI's predictions, and rate the accuracy of the prediction. Participants noted instances of inaccuracy when their emotion self-report contradicted the algorithm's prediction (Findings 5.2). In this way, by asking participants to note down instances of algorithmic error, our study engaged participants in an informal, exploratory audit.

Through this, our work draws on reflective design to engage calls by Costanza-Chock, Raji, and Buolamwini [36], and by Vecchione, Barocas, and Levy [147] for more participatory algorithmic accountability. Our probe offered a form of informal participatory algorithm auditing, leveraging the "power of everyday users in surfacing harmful algorithmic behaviors" [128, p. 1]. Our study suggests the potential of informal, participatory algorithm audits, different from and complementary to formal audits that systematically test algorithms. Through being open-ended, exploratory, and qualitative, informal participatory algorithm auditing can identify issues that technical experts may not have known to search for, and can more inclusively listen to diverse responses to algorithms that may not 'fit' within a more systematic approach. Note, we do not offer a fully fledged method for doing algorithm audits inspired by reflective design. Rather, we offer considerations for how reflective design can inform future work on informal, participatory auditing by facilitating participants' critical reflections on an algorithm for finding algorithmic errors.

We argue that our study fostered a kind of informal, participatory algorithm audit. We acknowledge that this differs significantly from more typical, formalized audits that systematically test results for a particular purpose.

6.2.1 Connecting our study to algorithm auditing. We unpack how what emerged in our study supports existing recommendations from prior auditing literature for participatory algorithm audits. Participants' critical reflections often touched on issues with Emotion AI that have already been found in algorithm audits, and/or potential issues with Emotion AI warranting further investigation through future algorithm audits. For example, participants identified inaccuracies related to lighting (Sec. 5.2.1), wearing glasses (5.2.2), wearing makeup (5.2.3), and axes of social discrimination (5.2.4). Overall, this illustrates that our probe and study surfaced insights of direct relevance to algorithm auditing. By suggesting potential factors contributing to inaccuracy that can suggest future algorithm audits, our project helped support "Deriving actionable

insights from user-engaged auditing," a challenge of participatory algorithm auditing identified by Deng et al. [40, Sec. 4.2.4].

Participants also offered broader reflections regarding fundamental critiques of Emotion AI (Findings 5.1) and patterns of miscategorization (Sec. 5.3). Future audits should investigate what contributes to these potentially stigmatizing patterns of miscategorization. In this way, our study supported "soliciting critical and holistic feedback from user auditors," a recommendation from Deng et al. on "effectively scaffolding users in auditing algorithms" [40, Section 4.2.3] as well as engage Deng et al.'s challenge of "focusing on 'who' " is doing the auditing [40, Section 6.1.1].

The post-interview offered a venue for sense-making for participants. Li et al. suggest "designing for discussion and deliberation" to support participants in auditing algorithms [91, Section 7.3.1]. DeVos et al. note the importance of collective sense-making where participants can discuss algorithmic results together [41]. Our work suggests that participant-researcher dialogue may also offer a venue supporting sense-making.

Drawing from reflective informatics [18], which builds on reflective design specifically for informatics, the probe invites moments of *breakdown* and *inquiry*, dimensions of reflective informatics. Based on our own informal use experiences with the Emotion AI, we anticipated that participant emotion self-report would often differ significantly from the Emotion AI predicted labels—these errors are moments of *breakdown* of the Emotion AI. The probe invites participants to *inquire* about the similarities or differences in how they describe their own emotions and how the Emotion AI categorizes their emotions. Intentionally exposing participants to Emotion AI errors also responds to Raji et al.'s point that too often AI ethics debates assume AI systems are functional [114].

Overall, our probe and study came out of our practices as design researchers who regularly engage reflective design, yet our work yielded findings relevant for algorithm auditing. In the next section, we zoom out from our own project to unpack considerations for how algorithm auditing might engage aspects of reflective design.

6.2.2 Connecting reflective design to algorithm auditing. We connect key principles of reflective design to novel recommendations for informal participatory algorithm auditing. How can future algorithm audits draw from reflective design to invite participants' holistic, critical feedback on algorithms, identifying algorithm errors, and deriving actionable insights? Returning to Sengers' et al. foundational paper defining reflective design [124], we unpack how principles of reflective design may help with this.

Reflective Design Principle 4: "Technology should support skepticism about and reinterpretation of its own working" [124, p. 55]. In designing the interface or system that participants will use for algorithm auditing, supporting skepticism can facilitate identifying errors. Supporting reinterpretation can invite more holistic critiques or feedback on the algorithm. In our case, we invited skepticism by asking participants to rate the accuracy of each prediction, which implies that sometimes a low accuracy rating might be appropriate. More broadly, this principle of reflective design suggests leaving open a space for participants to reject the technology's recommendation, and to position participants as the final authority on the subject. **For algorithm auditing, this suggests designing to invite skepticism about the algorithm being audited, and positioning participants as the authority.**

Reflective Design Principle 5: "Reflection is not a separate activity from action but is folded into it as an integral part of experience" [124, p. 56]. This suggests folding auditing into broader use, such as integrating features for participants to easily flag potential errors as they encounter them throughout their everyday use of an algorithmic system.

Reflective Design Principle 6: "Dialogic engagement between designers and users through technology can enhance reflection" [124, p. 56]. In our project, the semi-structured post-interview offered a dialogic engagement between designers and users. More broadly, this reflective design principle suggests that **exploratory, informal, participatory algorithm audits can create spaces for dialogue between researchers and participants to enhance critical reflections that can yield insights regarding the algorithm in question.**

Further, reflective design strategies, "Give users license to participate," "Provide dynamic feedback to users," and "Inspire rich feedback from users" [124, p. 56] could offer suggestions for inviting active participation, prompting continual re-evaluation of an algorithm through dynamic feedback, and gathering rich holistic feedback from participants.

Our project offers one example of leveraging reflective design to invite critical reflections from participants, yielding insights relevant for Emotion AI algorithm auditing. Here, we have stitched connections between reflective design and algorithm auditing more broadly. Future work should continue to explore how reflective design may offer useful considerations to inform the design of participatory, informal algorithm audits.

6.2.3 Position participant voices as ground truth. When seeking instances of algorithmic error or bias in auditing, how do users make sense of what is an error and what is biased? DeVos et al. found that one strategy is comparing the algorithm's results to what participants consider reality or fact [41, Section 4.5.2]. Our probe asked participants to first report their own emotions in their own words, providing a 'reality' 'ground truth' of their emotions, and then to examine Emotion AI's results, enabling this comparison. Reflective design calls for making the user "the final authority on what the user is doing" [124, p.55-56], or in our case feeling.

For Emotion AI, we treat participants' self-report of their own emotions as the ground truth, for ethical reasons. This contrasts an epistemological perspective often adopted in affective computing and Emotion AI, as exemplified by for instance Kaur et al. who suggest that participant self-report is not really a ground truth because participants may forget or omit emotions from their self-report, or may simply be unsure of how they are feeling [79]. We acknowledge that human experience and emotion are complex and multifaceted, and participant emotion self-report of emotions has flaws (e.g., [16, 132]). We have observed issues with participant emotion self-report in our own work. Despite this, we argue that participants' emotion self-report *must* be treated as ground truth, or the closest available proxy for ground truth, for ethical reasons, in order to **respect and preserve subjective agency for participants**. As one participant said, *"I think to get someone's emotional pain and suffering, we should just ask them. And then we should*

trust them, what they're telling." Similarly for positive emotions, participants must decide for themselves what feels good to them.

People's self-narration of their emotions, experiences, intentions, actions, etc., is essential to how people make meaning of their lives, and it is essential to respect and preserve participants' agency. Of course, participant's self-report could be revisited or revised over time through reflection, as Kaur et al. [79] observed when participants remembered and added to what they reported feeling—supporting Kaur et al.'s suggestion around enabling participants to annotate or edit their emotion data [79]. Participant self-report could be combined with approaches such as cued recall [30]. **We argue that, for auditing Emotion AI algorithms, the ground truth against which to evaluate algorithmic accuracy must be participant self-report of their own emotions.**

Beyond Emotion AI, how might participants using an algorithm access a 'ground truth' against which to evaluate the algorithms outputs? In some cases, the algorithm may be attempting to replicate a task with clear answers that most people can identify and agree on, such as image recognition of commonplace objects (e.g., [61, 128]). In other cases, a 'ground truth' could be difficult to establish due to the complexity or subjectivity of the assessment. This could especially relevant when when considering characteristics that are often unobserved by algorithms or impossible to quantify [140]. We suggest allowing participants to put forth their own sense of the 'ground truth' against which to compare the algorithm's output. The 'ground truth' assessment of multiple participants can be discussed, foregrounding voices of participants positioned to be most impacted by the algorithm. This can bring algorithm auditing closer to its historical roots in audits in the social sciences, which uncovered, for instance, racist hiring and housing discriminatory practices [97]. **For auditing beyond Emotion AI, we suggest opportunities for designers to solicit participants' thoughtful, reflective perceptions of their own 'ground truths' against which to evaluate an algorithm.**

7 LIMITATIONS & FUTURE WORK

Our probe explored one Emotion AI system, Morphcast [3], chosen for its privacy policy protections. Algorithms change over time as new versions are deployed, and other Emotion AI systems may perform better, worse, or differently. Finding particular errors with one version of one Emotion AI algorithm does not mean that other Emotion AI algorithms will exhibit the same errors; finding errors with one algorithm and connecting these to similar errors found in other algorithms could be one way of learning about broader patterns in algorithmic errors. Our study reached participants from different cultural backgrounds in the US, but further studies are needed evaluating Emotion AI in different cultures and languages.

Our work does not offer a fully-fledged method for integrating reflective design into algorithmic auditing. Our work shows rich insights from small-N qualitative approaches, and does not address how to 'scale up' to large-N studies often valued for auditing. Future work should continue exploring generative connections between critically-oriented design research approaches and informal participatory algorithm auditing.

8 CONCLUSION

This paper connects reflective design, Emotion AI, and algorithm audits. We initially intended to do a reflective design study to explore participants' critical reflections of Emotion AI. Algorithm auditing emerged later (Sec. 3.2) when we found that participants' responses identified numerous algorithm errors. By analyzing our study through the lens of informal participatory algorithm auditing, we uncovered connections between foundational reflective design principles and the aims of participatory auditing (Sec. 6.2). We contribute (Sec. 6.1) insights for designing Emotion AI, and (Sec. 6.2) considerations for how reflective design can inform informal, participatory, exploratory algorithm audits. Through this, this paper offers generative pathways for design research engagements with AI ethics.

ACKNOWLEDGMENTS

This work was supported by a Google TensorFlow faculty research award, a Georgia Tech Ivan Allen College Small Grants for Research award, and a Georgia Tech COVID19 Faculty Relief Program award. This material is based upon work supported by the National Science Foundation under Grant No. (2335974).

REFERENCES

[1] [n. d.]. Affectiva: Humanizing Technology. https://www.affectiva.com/
[2] [n. d.]. Department of Homeland Security Future Attribute Screening Technology Mobile Module (FAST M2) Overview. https://publicintelligence.net/dhs-future-attribute-screening-technology-mobile-module-fast-m2-overview/
[3] [n. d.]. Emotion AI Interactive Video Platform. https://www.morphcast.com/
[4] [n. d.]. Feel: Biomarkers & Digital Therapeutics for Mental Health. https://www.myfeel.co/
[5] [n. d.]. Feel: Biomarkers & Digital Therapeutics for Mental Health. https://www.myfeel.co/
[6] [n. d.]. HireVue. https://www.hirevue.com/products/video-interviewing
[7] [n. d.]. iMotions Facial Expression Analysis. https://imotions.com/biosensor/fea-facial-expression-analysis/
[8] [n. d.]. Limbic.AI is enabling the best psychological therapy. Our AI-powered patient reporting tool makes it easier than ever for patients to provide their clinician with the data they need to treat optimally. https://limbic-website.netlify.app//
[9] [n. d.]. Sound Intelligence Aggression Detection. https://www.soundintel.com/products/overview/aggression/
[10] [n. d.]. VCV Online Recruitment Automation Software: Virtual Staffing Solutions. https://vcv.ai
[11] [n. d.]. Ōura Ring: Accurate Health Information Accessible to Everyone. https://ouraring.com
[12] 2021. Bellabeat: Sync your body and mind. https://bellabeat.com/
[13] 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*. ACM, Honolulu HI USA, 1–16. https://doi.org/10.1145/3313831.3376680
[14] Joshua Attenberg, Panos Ipeirotis, and Foster Provost. 2015. Beat the Machine: Challenging Humans to Find a Predictive Model's "Unknown Unknowns". *Journal of Data and Information Quality* 6, 1 (March 2015), 1:1–1:17. https://doi.org/10.1145/2700832
[15] Jeffrey Bardzell and Shaowen Bardzell. 2013. What Is "Critical" about Critical Design?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3297–3306. https://doi.org/10.1145/2470654.2466451
[16] Lisa Feldman Barrett. 2004. Feelings or Words? Understanding the Content in Self-Report Ratings of Experienced Emotion. *Journal of Personality and Social Psychology* 87, 2 (Aug. 2004), 266–281. https://doi.org/10.1037/0022-3514.87.2.266
[17] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. 2019. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest* 20, 1 (July 2019), 1–68. https://doi.org/10.1177/1529100619832930
[18] Eric P.S. Baumer. 2015. Reflective Informatics: Conceptual Dimensions for Designing Technologies of Reflection. In *Proceedings of the 33rd Annual ACM*

Conference on Human Factors in Computing Systems (CHI '15). Association for Computing Machinery, New York, NY, USA, 585–594. https://doi.org/10.1145/2702123.2702234

[19] Eric P.S. Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. 2014. Reviewing Reflection: On the Use of Reflection in Interactive System Design. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. Association for Computing Machinery, New York, NY, USA, 93–102. https://doi.org/10.1145/2598510.2598598

[20] Eric P.S. Baumer and M. Six Silberman. 2011. When the Implication Is Not to Design (Technology). In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 2271. https://doi.org/10.1145/1978942.1979275

[21] Julie Beck. 2015. Hard Feelings: Science's Struggle to Define Emotions. https://www.theatlantic.com/health/archive/2015/02/hard-feelings-sciences-struggle-to-define-emotions/385711/

[22] Jake Bittle. 2020. Lie detectors have always been suspect. AI has made the problem worse. *MIT Technology Review* (March 2020). https://www.technologyreview.com/2020/03/13/905323/ai-lie-detectors-polygraph-silent-talker-iborderctrl-converus-neuroid/

[23] Kirsten Boehner. 2009. Reflections on representation as response. *interactions* 16, 6 (Nov. 2009), 28. https://doi.org/10.1145/1620693.1620700

[24] Kirsten Boehner, Rogério DePaula, Paul Dourish, and Phoebe Sengers. 2005. Affect: from information to interaction. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility (CC'05)*. ACM Press, 59–68. https://doi.org/10.1145/1094562.1094570

[25] Kirsten Boehner, Rogério DePaula, Paul Dourish, and Phoebe Sengers. 2007. How emotion is made and measured. *International Journal of Human-Computer Studies* 65, 4 (April 2007), 275–291. https://doi.org/10.1016/j.ijhcs.2006.11.016

[26] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI Interprets the Probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/1240624.1240789

[27] Esther Bosch, David Bethge, Marie Klosterkamp, and Thomas Kosch. 2022. Empathic Technologies Shaping Innovative Interaction: Future Directions of Affective Computing. In *Adjunct Proceedings of the 2022 Nordic Human-Computer Interaction Conference (NordiCHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–3. https://doi.org/10.1145/3547522.3547703

[28] Karen L. Boyd and Nazanin Andalibi. 2023. Automated Emotion Recognition in the Workplace: How Proposed Technologies Reveal Potential Futures of Work. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 95:1–95:37. https://doi.org/10.1145/3579528

[29] Simone Browne. 2018. B®anding Blackness - Biometric Technology and the Surveillance of Blackness. In *Sondra Perry: Typhoon Coming On*, Sondra Perry and Amira Gad (Eds.). Walther König, Köln.

[30] Anders Bruun, Effie Lai-Chong Law, Matthias Heintz, and Poul Svante Eriksen. 2016. Asserting Real-Time Emotions through Cued-Recall: Is It Valid?. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/2971485.2971516

[31] J. Buolamwini and T. Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of Machine Learning Research*, Vol. 81. 77–91.

[32] Mike Butcher. [n. d.]. The robot-recruiter is coming — VCV's AI will read your face in a job interview. http://social.techcrunch.com/2019/04/23/the-robot-recruiter-is-coming-vcvs-ai-will-read-your-face-in-a-job-interview/

[33] Rafael Calvo, Sidney D'Mello, Jonathan Gratch, and Arvid Kappas (Eds.). 2015. *The Oxford Handbook of Affective Computing*. Oxford University Press. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199942237.001.0001/oxfordhb-9780199942237

[34] Angela Chen and Karen Hao. 2020. Emotion AI researchers say overblown claims give their work a bad name. A lack of government regulation isn't just bad for consumers. It's bad for the field, too. *MIT Technology Review* (Feb. 2020). https://www.technologyreview.com/2020/02/14/844765/ai-emotion-recognition-affective-computing-hirevue-regulation-ethics/

[35] Shanley Corvite, Kat Roemmich, Tillie Ilana Rosenberg, and Nazanin Andalibi. 2023. Data Subjects' Perspectives on Emotion Artificial Intelligence Use in the Workplace: A Relational Ethics Lens. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 124:1–124:38. https://doi.org/10.1145/3579600

[36] Sasha Costanza-Chock, Inioluwa Deborah Raji, and Joy Buolamwini. 2022. Who Audits the Auditors? Recommendations from a Field Scan of the Algorithmic Auditing Ecosystem. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1571–1583. https://doi.org/10.1145/3531146.3533213

[37] P. Coulton, J. Lindley, M. Sturdee, and M. Stead. 2017. Design Fiction as World Building. In *Proceedings of the 3nd Biennial Research Through Design Conference*.

[38] Kate Crawford, Roel Dobbe, Theodora Dryer, Genevieve Fried, Ben Green, Elizabeth Kaziunas, Amba Kak, Varoon Mathur, Erin McElroy, Andrea Nill Sánchez, Deborah Raji, Joy Lisi Rankin, Rashida Richardson, Jason Schultz, Sarah Myers West, and Meredith Whittaker. 2019. *AI Now 2019 Report*. Technical Report. AI Now Institute, New York. 100 pages. https://ainowinstitute.org/AI_Now_2019_Report.pdf

[39] Max T. Curran, Jeremy Raboff Gordon, Lily Lin, Priyashri Kamlesh Sridhar, and John Chuang. 2019. Understanding Digitally-Mediated Empathy: An Exploration of Visual, Narrative, and Biosensory Informational Cues. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300844

[40] Wesley Hanwen Deng, Boyuan Guo, Alicia Devrio, Hong Shen, Motahhare Eslami, and Kenneth Holstein. 2023. Understanding Practices, Challenges, and Opportunities for User-Engaged Algorithm Auditing in Industry Practice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3544548.3581026

[41] Alicia DeVos, Aditi Dhabalia, Hong Shen, Kenneth Holstein, and Motahhare Eslami. 2022. Toward User-Driven Algorithm Auditing: Investigating Users' Strategies for Uncovering Harmful Algorithmic Behavior. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–19. https://doi.org/10.1145/3491102.3517441

[42] Nathalie DiBerardino and Luke Stark. 2023. (Anti)-Intentional Harms: The Conceptual Pitfalls of Emotion AI in Education. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1386–1395. https://doi.org/10.1145/3593013.3594088

[43] Anthony Dunne and Fiona Raby. 2001. *Design Noir: The Secret Life of Electronic Objects*. Birkhäuser, Basel.

[44] Anthony Dunne and Fiona Raby. 2013. *Speculative Everything: Design, Fiction, and Social Dreaming*. The MIT Press, Cambridge, Massachusetts ; London.

[45] Nabil Echchaibi. 2022. Mulsims between Transparency and Opacity. In *Cyber Muslims: mapping Islamic digital media in the internet age* (1 ed.), Robert Thomas Rozehnal (Ed.). Bloomsbury Academic, New York, 237–250.

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

[47] Paul Ekman and Erika L. Rosenberg (Eds.). 2005. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* (2nd ed ed.). Oxford University Press, Oxford ; New York.

[48] Severin Engelmann, Chiara Ullstein, Orestis Papakyriakopoulos, and Jens Grossklags. 2022. What People Think AI Should Infer From Faces. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 128–141. https://doi.org/10.1145/3531146.3533080

[49] Mohamed Ez-Zaouia, Aurélien Tabard, and Elise Lavoué. 2020. EMODASH: A dashboard supporting retrospective awareness of emotions in online learning. *International Journal of Human-Computer Studies* 139 (2020), 102411.

[50] Yingruo Fan, Jacqueline C. K. Lam, and Victor O. K. Li. 2021. Demographic Effects on Facial Emotion Expression: An Interdisciplinary Investigation of the Facial Action Units of Happiness. *Scientific Reports* 11 (March 2021), 5214. https://doi.org/10.1038/s41598-021-84632-9

[51] Jackson Feijó Filho, Thiago Valle, and Wilson Prata. 2014. Exploring Non-Verbal Communications in Mobile Text Chat: Emotion-Enhanced Chat. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14)*. Association for Computing Machinery, New York, NY, USA, 1069–1072. https://doi.org/10.1145/2639189.2670278

[52] Nick J Fox and Pam Alldred. 2023. Applied Research, Diffractive Methodology, and the Research-Assemblage: Challenges and Opportunities. *Sociological Research Online* 28, 1 (March 2023), 93–109. https://doi.org/10.1177/13607804211029978

[53] Christopher Frauenberger. 2019. Entanglement HCI The Next Wave? *ACM Transactions on Computer-Human Interaction* 27, 1 (Nov. 2019), 2:1–2:27. https://doi.org/10.1145/3364998

[54] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: Cultural Probes. *interactions* 6, 1 (Jan. 1999), 21–29. https://doi.org/10.1145/291224.291235

[55] William Gaver, Peter Gall Krogh, Andy Boucher, and David Chatting. 2022. Emergence as a Feature of Practice-based Design Research. In *Designing Interactive Systems Conference (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 517–526. https://doi.org/10.1145/3532106.3533524

[56] 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)*. ACM, New York, NY, USA, 233–240. https://doi.org/10.1145/642611.642653

[57] William W. Gaver, Andrew Boucher, Sarah Pennington, and Brendan Walker. 2004. Cultural Probes and the Value of Uncertainty. *Interactions* 11, 5 (Sept. 2004), 53–56. https://doi.org/10.1145/1015530.1015555

[58] Evelien Geerts and Iris van der Tuin. 2016. Diffraction & Reading Diffractively. https://newmaterialism.eu/almanac/d/diffraction.html

[59] Jack Gillum and Jeff Kao. 2019. Aggression Detectors: The Unproven, Invasive Surveillance Technology Schools Are Using to Monitor Students. *ProPublica* (June 2019). https://features.propublica.org/aggression-detector/the-unproven-invasive-surveillance-technology-schools-are-using-to-monitor-students/

[60] Gabriel Grill and Nazanin Andalibi. 2022. Attitudes and Folk Theories of Data Subjects on Transparency and Accuracy in Emotion Recognition. *Proceedings of the ACM on Human-Computer Interaction* (April 2022). https://doi.org/10.1145/3512925

[61] Jessica Guynn. 2015. Google Photos Labeled Black People 'Gorillas'. *USA TODAY* (July 2015). https://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/

[62] Bingyi Han, George Buchanan, and Dana Mckay. 2022. Learning in the Panopticon: Examining the Potential Impacts of AI Monitoring on Students. In *Proceedings of the 34th Australian Conference on Human-Computer Interaction*. ACM, Canberra ACT Australia, 9–21. https://doi.org/10.1145/3572921.3572937

[63] Drew Harwell. 2019. Rights group files federal complaint against AI-hiring firm HireVue, citing 'unfair and deceptive' practices. *Washington Post* (Nov. 2019). https://www.washingtonpost.com/technology/2019/11/06/prominent-rights-group-files-federal-complaint-against-ai-hiring-firm-hirevue-citing-unfair-deceptive-practices/

[64] Douglas Heaven. 2020. Why Faces Don't Always Tell the Truth about Feelings. *Nature* 578, 7796 (Feb. 2020), 502–504. https://doi.org/10.1038/d41586-020-00507-5

[65] Melissa Heikkilä. 2023. Five Things You Need to Know about the EU's New AI Act. https://www.technologyreview.com/2023/12/11/1084942/five-things-you-need-to-know-about-the-eus-new-ai-act/

[66] Alex Hern. 2020. Twitter Apologises for 'racist' Image-Cropping Algorithm. *The Guardian* (Sept. 2020). https://www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm

[67] Javier Hernandez, Rob R. Morris, and Rosalind Picard. 2011. Call Center Stress Recognition with Person-specific Models. In *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (ACII'11, Vol. 1)*. Springer-Verlag, Berlin, Heidelberg, 125–134. http://dl.acm.org/citation.cfm?id=2062780.2062798

[68] 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 (DIS '17)*. Association for Computing Machinery, New York, NY, USA, 95–99. https://doi.org/10.1145/3064663.3064703

[69] Tad Hirsch, Christina Soma, Kritzia Merced, Patty Kuo, Aaron Dembe, Derek D. Caperton, David C. Atkins, and Zac E. Imel. 2018. "It's Hard to Argue with a Computer": Investigating Psychotherapists' Attitudes towards Automated Evaluation. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. Association for Computing Machinery, New York, NY, USA, 559–571. https://doi.org/10.1145/3196709.3196776

[70] Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary P Czerwinski. 2022. Design of Digital Workplace Stress-Reduction Intervention Systems: Effects of Intervention Type and Timing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3491102.3502027

[71] Noura Howell, John Chuang, Abigail De Kosnik, Greg Niemeyer, and Kimiko Ryokai. 2018. Emotional Biosensing: Exploring Critical Alternatives. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 69:1–69:25. https://doi.org/10.1145/3274338

[72] Noura Howell, Laura Devendorf, Rundong (Kevin) Tian, Tomás Vega Gálvez, Nan-Wei Gong, Ivan Poupyrev, Eric Paulos, and Kimiko Ryokai. 2016. Biosignals as Social Cues: Ambiguity and Emotional Interpretation in Social Displays of Skin Conductance. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. Association for Computing Machinery, New York, NY, USA, 865–870. https://doi.org/10.1145/2901790.2901850

[73] Noura Howell, Laura Devendorf, Tomás Alfonso Vega Gálvez, Rundong Tian, and Kimiko Ryokai. 2018. Tensions of Data-Driven Reflection: A Case Study of Real-Time Emotional Biosensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174005

[74] Noura Howell, Greg Niemeyer, and Kimiko Ryokai. 2019. Life-Affirming Biosensing in Public: Sounding Heartbeats on a Red Bench. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3290605.3300910

[75] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology Probes: Inspiring Design for and with Families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 17–24. https://doi.org/10.1145/642611.642616

[76] Louise Marie Jupe and David Adam Keatley. 2020. Airport Artificial Intelligence Can Detect Deception: Or Am i Lying? *Security Journal* 33, 4 (Dec. 2020), 622–635. https://doi.org/10.1057/s41284-019-00204-7

[77] Marie Louise Juul Søndergaard, Nadia Campo Woytuk, Noura Howell, Vasiliki Tsaknaki, Karey Helms, Tom Jenkins, and Pedro Sanches. 2023. Fabulation as an Approach for Design Futuring. In *Designing Interactive Systems*.

[78] Edward B. Kang. 2023. On the Praxes and Politics of AI Speech Emotion Recognition. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 455–466. https://doi.org/10.1145/3593013.3594011

[79] Harmanpreet Kaur, Daniel McDuff, Alex C. Williams, Jaime Teevan, and Shamsi T. Iqbal. 2022. "I Didn't Know I Looked Angry": Characterizing Observed Emotion and Reported Affect at Work. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3491102.3517453

[80] Harmanpreet Kaur, Alex C. Williams, Daniel McDuff, Mary Czerwinski, Jaime Teevan, and Shamsi T. Iqbal. 2020. Optimizing for Happiness and Productivity: Modeling Opportune Moments for Transitions and Breaks at Work. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376817

[81] Joseph 'Jofish' Kaye. 2006. I Just Clicked to Say I Love You: Rich Evaluations of Minimal Communication. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. Association for Computing Machinery, New York, NY, USA, 363–368. https://doi.org/10.1145/1125451.1125530

[82] Os Keyes, Jevan Hutson, and Meredith Durbin. 2019. A Mulching Proposal: Analysing and Improving an Algorithmic System for Turning the Elderly into High-Nutrient Slurry. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290607.3310433

[83] Douwe Kiela, Max Bartolo, Yixin Nie, Divyansh Kaushik, Atticus Geiger, Zhengxuan Wu, Bertie Vidgen, Grusha Prasad, Amanpreet Singh, Pratik Ringshia, Zhiyi Ma, Tristan Thrush, Sebastian Riedel, Zeerak Waseem, Pontus Stenetorp, Robin Jia, Mohit Bansal, Christopher Potts, and Adina Williams. 2021. Dynabench: Rethinking Benchmarking in NLP. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Online, 4110–4124. https://doi.org/10.18653/v1/2021.naacl-main.324

[84] Goda Klumbytė, Phillip Lücking, and Claude Draude. 2020. Reframing AX with Critical Design: The Potentials and Limits of Algorithmic Experience as a Critical Design Concept. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3419249.3420120

[85] Sandjar Kozubaev, Chris Elsden, Noura Howell, Marie Louise Juul Søndergaard, Nick Merrill, Britta Schulte, and Richmond Y. Wong. 2020. Expanding Modes of Reflection in Design Futuring. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–15. https://doi.org/10.1145/3313831.3376526

[86] Kate Krosschell. [n. d.]. Facial Expression Analysis: The Complete Pocket Guide. https://imotions.com/blog/facial-expression-analysis/

[87] Krzysztof Kutt, Piotr Sobczyk, and Grzegorz J. Nalepa. 2022. Evaluation of Selected APIs for Emotion Recognition from Facial Expressions. In *Bio-Inspired Systems and Applications: From Robotics to Ambient Intelligence (Lecture Notes in Computer Science)*, José Manuel Ferrández Vicente, José Ramón Álvarez-Sánchez, Félix de la Paz López, and Hojjat Adeli (Eds.). Springer International Publishing, Cham, 65–74. https://doi.org/10.1007/978-3-031-06527-9_7

[88] Stacey Kuznetsov and Eric Paulos. 2010. Participatory Sensing in Public Spaces: Activating Urban Surfaces with Sensor Probes. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems (DIS '10)*. ACM, New York, NY, USA, 21–30. https://doi.org/10.1145/1858171.1858175

[89] Michelle S. Lam, Mitchell L. Gordon, Danaë Metaxa, Jeffrey T. Hancock, James A. Landay, and Michael S. Bernstein. 2022. End-User Audits: A System Empowering Communities to Lead Large-Scale Investigations of Harmful Algorithmic Behavior. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 512:1–512:34. https://doi.org/10.1145/3555625

[90] Agnieszka Landowska, Aleksandra Karpus, Teresa Zawadzka, Ben Robins, Duygun Erol Barkana, Hatice Kose, Tatjana Zorcec, and Nicholas Cummins. 2022. Automatic Emotion Recognition in Children with Autism: A Systematic Literature Review. *Sensors (Basel, Switzerland)* 22, 4 (Feb. 2022), 1649. https://doi.org/10.3390/s22041649

[91] Rena Li, Sara Kingsley, Chelsea Fan, Proteeti Sinha, Nora Wai, Jaimie Lee, Hong Shen, Motahhare Eslami, and Jason Hong. 2023. Participation and Division of Labor in User-Driven Algorithm Audits: How Do Everyday Users Work Together to Surface Algorithmic Harms?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3544548.3582074

[92] Fannie Liu, Geoff Kaufman, and Laura Dabbish. 2019. The Effect of Expressive Biosignals on Empathy and Closeness for a Stigmatized Group Member. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (Nov. 2019), 201:1–201:17. https://doi.org/10.1145/3359303

[93] Vidushi Marda and Shazeda Ahmed. 2021. *Emotional Entanglement: China's emotion recognition market and its implications for human rights.* Technical Report. Article 19. https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf

[94] Andrew McStay. 2020. Emotional AI and EdTech: serving the public good? *Learning, Media and Technology* 45, 3 (July 2020), 270–283. https://doi.org/10.1080/17439884.2020.1686016 Publisher: Routledge _eprint: https://doi.org/10.1080/17439884.2020.1686016.

[95] Andrew McStay. 2020. Emotional AI, Soft Biometrics and the Surveillance of Emotional Life: An Unusual Consensus on Privacy. *Big Data & Society* 7, 1 (Jan. 2020), 2053951720904386. https://doi.org/10.1177/2053951720904386

[96] Andrew McStay and P. Pavliscak. 2019. *Emotional Artificial Intelligence: Guidelines for Ethical Use.* Technical Report. https://drive.google.com/file/d/1frAGcvCY_v25V8ylqgPF2brTK9UVj_5Z/view

[97] Danaë Metaxa, Joon Sung Park, Ronald E. Robertson, Karrie Karahalios, Christo Wilson, Jeff Hancock, and Christian Sandvig. 2021. Auditing Algorithms: Understanding Algorithmic Systems from the Outside In. *Foundations and Trends® in Human–Computer Interaction* 14, 4 (2021), 272–344. https://doi.org/10.1561/1100000083

[98] Sebastian C. Müller and Thomas Fritz. 2015. Stuck and Frustrated or in Flow and Happy: Sensing Developers' Emotions and Progress. In *Proceedings of the 37th International Conference on Software Engineering (ICSE '15, Vol. 1).* IEEE Press, Piscataway, NJ, USA, 688–699. http://dl.acm.org/citation.cfm?id=2818754.2818838

[99] Lisa O'Carroll. 2023. EU Moves Closer to Passing One of World's First Laws Governing AI. *The Guardian* (June 2023). https://www.theguardian.com/technology/2023/jun/14/eu-moves-closer-to-passing-one-of-worlds-first-laws-governing-ai

[100] Rodrigo Ochigame and Katherine Ye. 2021. Search Atlas: Visualizing Divergent Search Results Across Geopolitical Borders. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference (DIS '21).* Association for Computing Machinery, New York, NY, USA, 1970–1983. https://doi.org/10.1145/3461778.3462032

[101] Jaspar Pahl, Ines Rieger, Anna Möller, Thomas Wittenberg, and Ute Schmid. 2022. Female, White, 27? Bias Evaluation on Data and Algorithms for Affect Recognition in Faces. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22).* Association for Computing Machinery, New York, NY, USA, 973–987. https://doi.org/10.1145/3531146.3533159

[102] Keyur Patel, Dev Mehta, Chinmay Mistry, Rajesh Gupta, Sudeep Tanwar, Neeraj Kumar, and Mamoun Alazab. 2020. Facial Sentiment Analysis Using AI Techniques: State-of-the-Art, Taxonomies, and Challenges. *IEEE Access* 8 (2020), 90495–90519. https://doi.org/10.1109/ACCESS.2020.2993803

[103] Eric Paulos and Tom Jenkins. 2005. Urban Probes: Encountering Our Emerging Urban Atmospheres. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05).* ACM, New York, NY, USA, 341–350. https://doi.org/10.1145/1054972.1055020

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

[105] Jesse Pepping, Sarah Scholte, Marnix van Wijland, Milan de Meij, Günter Wallner, and Regina Bernhaupt. 2020. Motiis: Fostering Parents' Awareness of Their Adolescents Emotional Experiences during Gaming. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20).* Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3419249.3420173

[106] Rosalind Picard. 1997. *Affective Computing.* The MIT Press.

[107] James Pierce, Phoebe Sengers, Tad Hirsch, Tom Jenkins, William Gaver, and Carl DiSalvo. 2015. Expanding and Refining Design and Criticality in HCI. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15).* Association for Computing Machinery, New York, NY, USA, 2083–2092. https://doi.org/10.1145/2702123.2702438

[108] Winifred Poster. 2019. Racialized Surveillance in the Digital Service Economy. In *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life*, Ruha Benjamin (Ed.). Duke University Press, Durham, 133–170. https://www.dukeupress.edu/captivating-technology

[109] Evani Radiya-Dixit and Gina Neff. 2023. A Sociotechnical Audit: Assessing Police Use of Facial Recognition. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23).* Association for Computing Machinery, New York, NY, USA, 1334–1346. https://doi.org/10.1145/3593013.3594084

[110] Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Jan. 2020), 469–481. https://doi.org/10.1145/3351095.3372828 arXiv: 1906.09208.

[111] Nina Rajcic and Jon McCormack. 2023. Message Ritual: A Posthuman Account of Living with Lamp. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23).* Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3544548.3581363

[112] Inioluwa Deborah Raji. 2022. From Algorithmic Audits to Actual Accountability: Overcoming Practical Roadblocks on the Path to Meaningful Audit Interventions for AI Governance. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22).* Association for Computing Machinery, New York, NY, USA, 5. https://doi.org/10.1145/3514094.3539566

[113] Inioluwa Deborah Raji and Joy Buolamwini. 2022. Actionable Auditing Revisited: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. *Commun. ACM* 66, 1 (Dec. 2022), 101–108. https://doi.org/10.1145/3571151

[114] Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst. 2022. The Fallacy of AI Functionality. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22).* Association for Computing Machinery, New York, NY, USA, 959–972. https://doi.org/10.1145/3531146.3533158

[115] Lauren Rhue. 2018. Racial Influence on Automated Perceptions of Emotions. *SSRN Electronic Journal* (2018). https://doi.org/10.2139/ssrn.3281765

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

[117] Kat Roemmich, Tillie Rosenberg, Serena Fan, and Nazanin Andalibi. 2023. Values in Emotion Artificial Intelligence Hiring Services: Technosolutions to Organizational Problems. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 109:1–109:28. https://doi.org/10.1145/3579543

[118] Kat Roemmich, Florian Schaub, and Nazanin Andalibi. 2023. Emotion AI at Work: Implications for Workplace Surveillance, Emotional Labor, and Emotional Privacy. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23).* Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3544548.3580950

[119] Camilo Rojas, Malena Corral, Niels Poulsen, and Pattie Maes. 2020. Project Us: A Wearable for Enhancing Empathy. In *Companion Publication of the 2020 ACM Designing Interactive Systems Conference.* ACM, Eindhoven Netherlands, 139–144. https://doi.org/10.1145/3393914.3395882

[120] Camilo Rojas, Eugenio Zuccarelli, Alexandra Chin, Gaurav Patekar, David Esquivel, and Pattie Maes. 2022. Towards Enhancing Empathy Through Emotion Augmented Remote Communication. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts.* ACM, New Orleans LA USA, 1–9. https://doi.org/10.1145/3491101.3519797

[121] Samiha Samrose, Daniel McDuff, Robert Sim, Jina Suh, Kael Rowan, Javier Hernandez, Sean Rintel, Kevin Moynihan, and Mary Czerwinski. 2021. MeetingCoach: An Intelligent Dashboard for Supporting Effective & Inclusive Meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21).* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445615

[122] Pedro Sanches, Noura Howell, Vasiliki Tsaknaki, Tom Jenkins, and Karey Helms. 2022. Diffraction-in-Action: Designerly Explorations of Agential Realism Through Lived Data. In *CHI Conference on Human Factors in Computing Systems (CHI '22).* Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3491102.3502029

[123] Pedro Sanches, Axel Janson, Pavel Karpashevich, Camille Nadal, Chengcheng Qu, Claudia Daudén Roquet, Muhammad Umair, Charles Windlin, Gavin Doherty, Kristina Höök, and Corina Sas. 2019. HCI and Affective Health: Taking Stock of a Decade of Studies and Charting Future Research Directions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3290605.3300475

[124] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. 2005. Reflective Design. In *Proceedings of the 4th Decennial Conference on Critical Computing (CC '05).* 49–58. https://doi.org/10.1145/1094562.1094569

[125] Phoebe Sengers, Kirsten Boehner, Michael Mateas, and Geri Gay. 2008. The disenchantment of affect. *Personal and Ubiquitous Computing* 12, 5 (June 2008), 347–358. https://doi.org/10.1007/s00779-007-0161-4

[126] Phoebe Sengers, Kirsten Boehner, Simeon Warner, and Tom Jenkins. 2005. Evaluating Affector: Co-Interpreting What "Works". In *CHI 2005 Workshop on Innovative Approaches to Evaluating Affective Interfaces.*

[127] Phoebe Sengers, Rainer Liesendahi, Werner Magar, Christoph Seibert, Boris Müller, Thorston Joachims, Weidong Geng, Pia M\a artensson, and Kristina Höök. 2002. The Enigmatics of Affect. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '02).* ACM, New York, NY, USA, 87–98. https://doi.org/10.1145/778712.778728

[128] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday Algorithm Auditing: Understanding the Power of Everyday Users in Surfacing Harmful Algorithmic Behaviors. *Proceedings of the ACM on*

Human-Computer Interaction 5, CSCW2 (Oct. 2021), 433:1–433:29. https://doi.org/10.1145/3479577

[129] Mona Sloane. 2021. The Algorithmic Auditing Trap. https://onezero.medium.com/the-algorithmic-auditing-trap-9a6f2d4d461d

[130] Petr Slovak, Alissa Antle, Nikki Theofanopoulou, Claudia Daudén Roquet, James Gross, and Katherine Isbister. 2023. Designing for Emotion Regulation Interventions: An Agenda for HCI Theory and Research. *ACM Transactions on Computer-Human Interaction* 30, 1 (March 2023), 13:1–13:51. https://doi.org/10.1145/3569898

[131] Petr Slovak, Alissa N. Antle, Nikki Theofanopoulou, Claudia Daudén Roquet, James J Gross, and Katherine Isbister. 2022. Designing for Emotion Regulation Interventions: An Agenda for HCI Theory and Research. (2022). https://doi.org/10.48550/ARXIV.2204.00118

[132] Katharine E Smidt and Michael K Suvak. 2015. A Brief, but Nuanced, Review of Emotional Granularity and Emotion Differentiation Research. *Current Opinion in Psychology* 3 (June 2015), 48–51. https://doi.org/10.1016/j.copsyc.2015.02.007

[133] Meredith Somers. 2019. Emotion AI, Explained.

[134] Andreas Sonderegger, Denis Lalanne, Luisa Bergholz, Fabien Ringeval, and Jürgen S. Sauer. 2013. Computer-Supported Work in Partially Distributed and Co-located Teams: The Influence of Mood Feedback. In *Human-Computer Interaction - INTERACT 2013 - 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 8118)*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Springer, 445–460. https://doi.org/10.1007/978-3-642-40480-1_30

[135] Luke Stark and Jesse Hoey. 2021. The Ethics of Emotion in Artificial Intelligence Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 782–793. https://doi.org/10.1145/3442188.3445939

[136] Luke Stark and Jevan Hutson. 2021. Physiognomic Artificial Intelligence. https://doi.org/10.2139/ssrn.3927300

[137] Ekaterina R. Stepanova, John Desnoyers-Stewart, Alexandra Kitson, Bernhard E. Riecke, Alissa N. Antle, Abdallah El Ali, Jeremy Frey, Vasiliki Tsaknaki, and Noura Howell. 2023. Designing with Biosignals: Challenges, Opportunities, and Future Directions for Integrating Physiological Signals in Human-Computer Interaction. In *Companion Publication of the 2023 ACM Designing Interactive Systems Conference (DIS '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 101–103. https://doi.org/10.1145/3563703.3591454

[138] Bruce M. Tharp and Stephanie M. Tharp. 2018. *Discursive Design: Critical, Speculative, and Alternative Things.* The MIT Press, Cambridge, MA.

[139] Stefan Timmermans and Iddo Tavory. 2012. Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis. *Sociological Theory* 30, 3 (Sept. 2012), 167–186. https://doi.org/10.1177/0735275112457914

[140] Nenad Tomasev, Kevin R. McKee, Jackie Kay, and Shakir Mohamed. 2021. Fairness for Unobserved Characteristics: Insights from Technological Impacts on Queer Communities. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)*. Association for Computing Machinery, New York, NY, USA, 254–265. https://doi.org/10.1145/3461702.3462540

[141] Timothy Torres. 2016. Pip Stress Management Biosensor Review. https://www.pcmag.com/reviews/pip-stress-management-biosensor

[142] Wenn-Chieh Tsai, Daniel Orth, and Elise van den Hoven. 2017. Designing Memory Probes to Inform Dialogue. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 889–901. https://doi.org/10.1145/3064663.3064791

[143] Vasiliki Tsaknaki, Tom Jenkins, Laurens Boer, Sarah Homewood, Noura Howell, and Pedro Sanches. 2020. Challenges and Opportunities for Designing with Biodata as Material. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–3. https://doi.org/10.1145/3419249.3420063

[144] Vasiliki Tsaknaki, Pedro Sanches, Tom Jenkins, Noura Howell, Laurens Boer, and Afroditi Bitzouni. 2022. Fabulating Biodata Futures for Living and Knowing Together. In *Designing Interactive Systems Conference (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1878–1892. https://doi.org/10.1145/3532106.3533477

[145] Terumi Umematsu, Akane Sano, and Rosalind Picard. 2019. Daytime Data and LSTM can Forecast Tomorrow's Stress, Health, and Happiness. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2186–2190. https://doi.org/10.1109/EMBC.2019.8856862 ISSN: 1558-4615.

[146] Kristen Vaccaro, Karrie Karahalios, Deirdre K. Mulligan, Daniel Kluttz, and Tad Hirsch. 2019. Contestability in Algorithmic Systems. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing (CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 523–527. https://doi.org/10.1145/3311957.3359435

[147] Briana Vecchione, Karen Levy, and Solon Barocas. 2021. Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '21)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3465416.3483294

[148] Neil Vigdor. 2019. Apple Card Investigated After Gender Discrimination Complaints. *The New York Times* (Nov. 2019). https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html

[149] James Vincent. 2019. AI "Emotion Recognition" Can't Be Trusted. https://www.theverge.com/2019/7/25/8929793/emotion-recognition-analysis-ai-machine-learning-facial-expression-review

[150] Ben Virdee-Chapman. [n. d.]. Kairos: IPG Case Study: Audience Analytics with Face Recognition. https://www.kairos.com/ipg

[151] Elizabeth Anne Watkins. 2023. Face Work: A Human-Centered Investigation into Facial Verification in Gig Work. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 1–24. https://doi.org/10.1145/3579485

[152] Daricia Wilkinson, Kate Crawford, Hanna Wallach, Deborah Raji, Bogdana Rakova, Ranjit Singh, Angelika Strohmayer, and Ethan Zuckerman. 2023. Accountability in Algorithmic Systems: From Principles to Practice. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3544549.3583747

A PARTICIPANT DEMOGRAPHICS

Participant demographics are summarized in the table on the following page.

P#	Pronouns	Self-described race(s) and/or ethnicity(ies)	Age	Before joining this study, what had you heard about AI and/or Emotion AI? What prior experience, if any, did you have with AI and/or Emotion AI?	Occupation
P1	she/her/hers	Asian	25-29		
P2	she/her/hers	White/ Caucasian	25-29	I've got a fair amount of AI experience (education AI research project for nearly three years), but my experience with emotion AI is pretty limited.	PhD Student
P3	he/him/his		18-24	Mainly, I heard about AI from my teammate and I took couple classes during college time.	Student
P4	she/her/hers	Asian	25-29	I've used the Affectiva AI once for a class.	PhD Student
P5	he/him/his	South-East Asian/Indian	30-39	I ran a company that produced training data for AI. My current company also builds some AI software.	Full time job
P6	she/her/hers	Asian	25-29	I have heard of AI a lot but have never heard of Emotion AI.	Full time job
P7	she/her/hers	Asian/White	30-39	I have heard of emotion AI, which I understand as using AI technologies to interpret/measure/understand human emotions (e.g. through facial recognition) but don't have much experience with AI otherwise.	PhD Student
P8	he/him/his	South Asian	18-24		
P9	she/her/hers	Black or African-American/ Nigerian	18-24	Previously, I knew that AI is the ability for machines or systems to think and behave similar to human intelligence and behavior. I knew little about emotion AI, but based on what I know about AI, I assumed that it had to do with understanding and reproducing human emotion. I briefly researched and discussed the ethics and implications of AI-related topics and implementations, including self-driving cars, robots, and algorithmic bias, in an undergraduate computing ethics seminar. I also did a research paper on AI in graduate school.	Full time job
P10	she/her/hers	Filipino, Spanish	18-24	I learned about AI through books and classes during my undergraduate career. After graduating, I'm volunteering with a group which enables me to look at AI Ethics and tech ethics.	Full time job
P11	she/her/hers	Chinese	18-24	Artificial intelligence almost mimics the human mind. The closest "AI" I've dealt with is the Siri	Student
P12	she/her/hers	Hispanic Latinx	18-24	I haven't really heard much but I know that the AI recognition is very biased and skewed to whyte folks because that is all the data was collected was conducted on. I would love to help diversify even just a little in your research in this specific aspect.	Student, part-time job
P13	he/him/his	Indian	25-29	I think AI is defined for a particular task where the algorithm or agent is able to perform human like activities	Student
P14	he/him/his	Asian/White	18-24	May be something like capturing facial expressions to understand human emotions. Have no prior experience	MS Student
P15	he/him/his	South Asian	18-24	No prior experience with Emotion AI. Have heard of but not worked with AI.	Student
P16	he/him/his	Asian/White	18-24	Been developing my own AI for simulation testing, was curious about this project applications	Student
P17	they/them/theirs	South Asian	18-24	None	Student
P18	he/him/his	White	18-24	I don't know a ton about it, as AI is not a huge interest for me. I do know it's a hot topic in computing concerning ethics and how conflicts should be handled	Student
P19	she/her/hers	white/asian	18-24	Very little, just that it can be helpful in some applications such as autonomous vehicles as well as creating things e.g. art. No experience.	Student
P20	he/him/his	White/Latino	30-39	I'm doing robotics, so a pretty good grasp on general AI concepts, but nothing about emotion AI	PhD Student
P21	he/him/his	Asian	25-29	I am aware to a certain extent about AI and the developments happening in the field, but not a lot of exposure.	Student
P22	she/her/hers	south asian	30-39	none	Student

Table 1: Participant demographics and prior experience with Emotion AI. Blank cells indicate that the participant left this response blank.