



The methodology of studying fairness perceptions in Artificial Intelligence: Contrasting CHI and FAccT

Niels van Berkel^{a,*}, Zhanna Sarsenbayeva^b, Jorge Goncalves^c

^a Aalborg University, Aalborg, Denmark

^b The University of Sydney, Sydney, Australia

^c University of Melbourne, Melbourne, Australia

ARTICLE INFO

Dataset link: <https://osf.io/7dfz5/>

Keywords:

Fairness
Equity
Algorithmic fairness
Machine learning
Artificial intelligence
Local standards
User studies
Algorithms

ABSTRACT

The topic of algorithmic fairness is of increasing importance to the Human-Computer Interaction research community following accumulating concerns regarding the use and deployment of Artificial Intelligence-based systems. How we conduct research on algorithmic fairness directly influences our inferences and conclusions regarding algorithmic fairness. To better understand the methodological decisions of studies focused on people's perceptions of algorithmic fairness, we systematically analysed relevant papers from the CHI and FAccT conferences. We identified 200 relevant papers published between 1993 and 2022 and assessed their study design, participant sample, and geographical location of participants and authors. Our results highlight that studies are predominantly cross-sectional, cover a wide range of participant roles, and that both authors and participants are primarily from the United States. Based on these findings, we reflect on the potential pitfalls and shortcomings in how the community studies algorithmic fairness.

1. Introduction

Increasing societal concerns surrounding the development, use, and deployment of Artificial Intelligence (AI) systems have led to a growing academic focus on the fairness of these AI systems. The inaugural Conference on Fairness, Accountability, and Transparency took place in 2018 and was affiliated with the ACM in 2019¹ and the ACM Conference on Computing and Sustainable Societies launched in 2020, indicating a growing understanding in the scientific Computer Science community of the need for ethical technological development. The fairness of algorithm-based systems, particularly AI systems, has become a widely debated topic within the Human-Computer Interaction (HCI) community (Abdul et al., 2018; Holstein et al., 2019; Woodruff et al., 2018; Alkhatib, 2021). Algorithmic fairness has been discussed in the context of the platform economy (Ma et al., 2018; Ahmed et al., 2016; Kleinberg et al., 2018), healthcare (Yang et al., 2016; Uhde et al., 2020), and social media (Alvarado and Waern, 2018; DeVito et al., 2017), among numerous others, highlighting the plethora of application areas in which algorithmic fairness is of increased concern.

This increasing urgency towards fair and trustworthy algorithms has resulted in a growing focus on the intended users and stakeholders of AI systems. Therefore, we set out to map and compare how publications at the HCI community's primary outlet (*ACM Conference on Human Factors*

in Computing Systems (CHI)) and a representative and growing community dealing with algorithmic fairness (*ACM Conference on Fairness, Accountability, and Transparency (FAccT)*) study the perceptions of end-users towards AI. In particular, this paper focuses on the mechanics and methods employed by researchers. The people we study, as well as the methodological choices underpinning our studies, severely impact the outcomes of our research and thereby drive the perspectives we can provide on algorithmic fairness. Prior reflections on methodological practices within other research communities have led researchers to identify weaknesses in existing methodological approaches, opening the way to future improvements within the discipline. One of the most widely known examples is the critique of study samples in Psychology. Arnett argued in 2008 that “by concentrating primarily on Americans, psychological researchers in the United States restrict their focus to less than 5% of the world’s total population. The rest of the world’s population, the other 95%, is neglected.” (Arnett, 2008, p. 602). This critique was later expanded upon by Henrich et al. in what has famously become known as ‘WEIRD study samples’ (Henrich et al., 2010). WEIRD (Western, Educated, Industrialised, Rich, and Democratic) participants are widely employed by researchers but are likely to differ in their behaviour, perceptions, and responses from a non-WEIRD study sample. A recent

* Corresponding author.

E-mail address: nielsvanberkel@cs.aau.dk (N. van Berkel).

¹ ACM FAccT, formerly known as ACM FAT* and the spiritual successor to FAT ML (<https://www.fatml.org/>).

study by Linxen et al. analysed the study sample of recent CHI papers and found similar results (Linxen et al., 2021), highlighting the dominance of Western participants in CHI studies.

Within the topic of algorithmic fairness, such differences can be of critical importance due to contrasting perspectives on algorithmic fairness and disparity in regulations related to data collection, storage, and usage between countries. A well-known example that highlights these differences in perspective is Awad et al.'s 'Moral Machine' study (Awad et al., 2018). In this study, participants were asked to select their preferred outcome in hypothetical autonomous driving accident scenarios. The geographically diverse participant sample allowed the authors to capture differences in preferences between geographical regions. A clustering analysis revealed a Western, Eastern, and Southern cluster—each prioritising different factors in determining their preferred outcome scenarios. Prior studies of governmental AI policy and strategy documents similarly highlight differences between countries' visions on (ethical) issues related to AI technology (van Berkel et al., 2020; Dexe and Franke, 2020).

In this paper, we systematically review and assess how the CHI and FAccT research communities have studied algorithmic fairness. We present a systematic assessment of 200 papers, following an initial identification of 1260 papers in the ACM Digital Library. Our analysis focuses on both the chosen study designs (e.g., study method), characteristics of the participant samples (e.g., sample size), as well as the geographical location of the studies' participants and authors. The results of our analysis highlight that the majority of studies rely on participants from the USA, are limited to single observations of participants (i.e., cross-sectional in nature), and recruit participants across a variety of roles (e.g., general public, domain experts). We provide implications for the future study of algorithmic fairness to decrease the existing gaps in geographical coverage, to increase our understanding of the effect of factors such as participant's role, their compensation, and their demographic characteristics, as well as to allow for increased comparability between studies on this growing topic.

2. Related work

In this section we introduce algorithmic fairness research within HCI and the wider Computer Science domain. Subsequently, we highlight earlier work assessing and discussing methodological concerns within a range of disciplines, followed by a closer look at the results from prior reflection on HCI research practices.

2.1. Algorithmic fairness

Fuelled by the rise of AI systems and the numerous reports on the harm caused by these systems (Rahim, 2020; Israni, 2017), the study of algorithmic fairness has become a critical component of AI research. Lepri et al. define algorithmic fairness as "*the lack of discrimination or bias in the decisions*" (Lepri et al., 2018, p. 615). Various works within Computer Science have aimed to formalise the notion of fairness, often distinguishing between different types of fairness, e.g. group fairness (Calders and Verwer, 2010) and individual fairness (Dwork et al., 2012). Whereas group fairness argues for an equal outcome of results between (demographic) groups (e.g., as based on race), individual fairness dictates that individuals with similar characteristics should be treated similarly. The increasing focus on algorithmic fairness has also led to a more critical and in-depth discussion of work towards algorithmic fairness.

A prominent example of the more critical discussion of algorithmic fairness is the work by Hoffmann, who highlights three limitations of the contemporary debate on algorithmic fairness (Hoffmann, 2019). First, Hoffmann states that algorithmic fairness copies the 'bad actor' frame, through which (un)fairness is reduced to narrow interpretations of 'cause and effect' by a bad actor who needs to be penalised. Such a frame leaves out important social and contextual issues. Second, work

on algorithmic fairness takes an insufficiently intersectional approach, instead focusing on an individual dimension of discrimination (e.g., race, gender), centred on disadvantages. An intersectionality approach argues that we cannot rely solely on pre-existing categories but instead should consider the interaction between dimensions of discrimination (e.g., gender and racial discrimination as experienced by Black women in employment discrimination (Crenshaw, 1989)), see also the foundational work by Crenshaw (1989). Third, Hoffmann argues that algorithmic fairness should not focus solely on achieving a fair distribution of goods (e.g., resources, opportunities), as an improvement in these metrics does not necessarily improve the day-to-day life of those facing discrimination (Hoffmann, 2019). Furthermore, we highlight the work by Verma and Rubin (2018), which discusses and compares twenty different definitions of fairness. Verma & Rubin conclude that "*the same case can be considered fair according to some definitions and unfair according to others*" (Verma and Rubin, 2018, p. 1). To overcome the adverse outcomes of algorithmic systems, Alkhatib states that "*designers should be designing to undermine these technologies*" (Alkhatib, 2021, p. 7). These examples of prior work highlight the inherent complexity of algorithmic fairness research, as well as the wide range of perspectives on algorithmic fairness in the field.

Within the HCI community, algorithmic fairness has been studied from a number of angles. A commonly recurring topic is the call for a broader view on algorithmic fairness, taking into consideration the various stakeholders involved (e.g., general public, policymakers, corporations) (DeVito et al., 2017; Saxena et al., 2019; Smith et al., 2020). Woodruff et al. present how workshops and interviews with members of traditionally marginalised populations can provide meaningful insights for developers and policymakers involved in the (potential) deployment of algorithmic systems (Woodruff et al., 2018). Non-experts have also been asked to comment on the fairness of individual variables/predictors that can be included in an algorithm, commonly through the use of crowdsourced data collection (e.g., van Berkel et al. (2019) and Grgic-Hlaca et al. (2018)). Input from such evaluations can be used to mark 'protected' variables or attributes. Marking a variable as protected requires that the decision of an algorithm is independent of the value of the protected variable (e.g., regardless of a person's age), increasing the fairness of prediction algorithms (Kleinberg et al., 2018).

Another branch of work focused on algorithmic fairness has investigated novel tools to support those involved in the development pipeline of algorithmic systems. As shown by prior work, human bias that emerges in this pipeline negatively affects the fairness of the resulting algorithm (Adam, 1998; Bowker and Star, 2000; Holstein et al., 2019). Brownstein defines implicit bias as "*a term of art referring to relatively unconscious and relatively automatic features of prejudiced judgment and social behavior.*" (Brownstein, 2019, p. 1). One such bias is the confirmation bias, in which we (often unconsciously) distort available information to confirm our pre-existing beliefs. Already in 1998, Adam discussed how male developers (unknowingly) embedded their own biases into AI applications (Adam, 1998). More recent examples point to the bias introduced by crowdworkers responsible for annotating training data on which the algorithm is subsequently trained (Attenberg et al., 2011; Vaughan, 2017). Several approaches have been proposed to counter such algorithmic biases, e.g., modification of the labels, addition of synthetic instances, and data transformation (Calders and Žliobaitė, 2013). The literature also focuses on tools that allow developers to interactively inspect the decisions made by an algorithm in hypothetical scenarios (Wexler et al., 2020).

As highlighted through these examples, the HCI literature on algorithmic fairness considers a variety of participant roles. In this work, we aim to map the various roles encountered in work on algorithmic fairness perceptions and the methodologies employed to answer our research questions.

2.2. Bias in research & practice

Internal reflection on the scientific practices of the global research community has been fundamental in the evolution of our scientific methods. Well-known contemporary examples include the replication crisis (Pashler and Wagenmakers, 2012), the over-reliance on participants from Western, Educated, Industrialised, Rich, and Democratic (WEIRD) countries (Henrich et al., 2010) (as recently studied in the context of HCI (Linxen et al., 2021)), *p*-value hacking (Dragicevic, 2016), and ethical and methodological concerns regarding the recruitment and compensation of crowdworkers (Irani and Silberman, 2013). The replication crisis, initially brought to light from within the social sciences and medicine, highlighted that many experimental results could not be reproduced through repeated experiments (Pashler and Wagenmakers, 2012). The inability to reproduce earlier results calls into question the reliability of prior studies and the methods and analysis techniques employed. Within HCI, replication of prior studies has been discussed primarily in light of publicly sharing study data and software and pre-registration of studies and analyses (Cockburn et al., 2018; Wacharamanatham et al., 2020). An analysis of 891 papers by Hornbæk found that 3% of papers were active replication studies, but also indicated that many of the non-replication studies could have included a replication of prior results with relatively minor additional effort (Hornbæk et al., 2014).

Within our own community, recent work has also discussed biases related to disabilities, gender, and race in technology. Spiel & Gerling analyse 66 publications on games aimed at neurodivergent populations (Spiel and Gerling, 2021). Their results highlight that current work is predominantly based on a medical model of disability and promotes a top-down approach to game development rather than proactive involvement of the target audience. The authors also call on future work to “attending to differences without articulating them as deficit” (Spiel and Gerling, 2021, p. 1). Highly relevant to this paper’s topic is Keyes’ work on ‘the misgendering machines’ (Keyes, 2018), which assesses how the HCI community operationalises gender and its use of ‘Automatic Gender Recognition’ systems. Keyes’ analysis highlights that these gender recognition systems consistently deny the existence of transgender people and the role of self-knowledge in relation to gender. Based on these findings, Keyes urges that HCI researchers need to both operationalise and understand gender in a nuanced way to put an end to the continued harm caused to trans people. This call for a more nuanced approach to gender has been repeated by Hamidi et al. (2018) and Scheuerman et al. (2019), among others. Schlesinger et al. raise the question of why it is challenging for chatbots to talk about race (Schlesinger et al., 2018). Their reflection identifies the widely used ‘ban list’ of prohibited words, aimed to reduce racist discussion, as a problematic factor as it limits and outright prevents race talk. Instead, Schlesinger et al. argue that chatbots can learn from real-world race talk conversations. Further, their work highlights the notion of context as an overlooked factor in chatbots (Schlesinger et al., 2018). Whereas the current generation of chatbots is assumed to respond appropriately across all possible contexts, the authors argue that domain-specific bots can perform more aptly in areas of conversation in which current bots’ behaviours are offensive or undesirable. Ogbonnaya-Ogburu et al. in discussing ‘critical race theory’, highlight how racism is abundant in our everyday socio-technical systems (Ogbonnaya-Ogburu et al., 2020). The authors state that, although the myriad of examples provided might be interpreted as an ‘aberration’ or ‘individual incident’ by Whites, they are part of daily reality for persons of colour.

These examples all point to biases that arise in research and practice, and the severe negative impact these biases can have on individuals that deviate from ‘the norm’. As an applied example from the context of Human-AI, Chancellor et al. studied how research on mental health discussion on social media has represented their study samples (Chancellor et al., 2019). Their results highlight a framing

by researchers towards the analysed social media users as ‘patients’ despite the lack of active and ongoing clinical care, as ‘participants’ even though these social media users often did not sign up for the study (automated scraping), as well as ‘social media users’ and ‘data/machine learning objects’ (in addition to the ‘humans’ framing). Such framing, Chancellor et al. argue, can risk dehumanising the individuals involved in the presented studies (Chancellor et al., 2019).

Biases may arise due to a wide array of factors. Friedman & Nissenbaum distinguish three categories of bias in computer systems, preexisting (*i.e.*, biases with roots in society), technical (*i.e.*, biases due to technical considerations), and emergent (*i.e.*, biases that arise during use of the system) (Friedman and Nissenbaum, 1996). Prior work in HCI has aimed to identify and quantify emergent biases in evaluation studies. In 1999, Nass et al. found that participants who had just completed a text-editing task were more positive in their interview answers when the interview took place on the same computer as compared to both their answers on a pen-and-paper questionnaire and their answers on an identical but different computer (Nass et al., 1999). The same effect is found with human interviewers, as shown by Dell et al. who report that participants’ bias towards an interviewer’s artefact can be influenced by the interviewer’s favour and demand characteristics (Dell et al., 2012). Participants’ evaluation of artefacts increased by 2.5x when participants believed that the interviewer developed the artefact, and by a factor of five when the researcher was a foreigner (requiring a translator). Vashistha et al. investigated the effects of social influence on participant response bias by showing positive, negative, or no feedback commentaries on videos that were subsequently rated by the participants (Vashistha et al., 2018). Their results indicate that participants presented with positive feedback videos provided high ratings and more positive feedback for the videos, whereas participants in the negative feedback condition generally provided low ratings and more critical feedback for the videos (Vashistha et al., 2018).

Similar to biases occurring at the individual study level, biases can also exist within research communities. By systematically assessing current methodological practices in the CHI and FAccT research community, we set out to map and contrast current practices in two communities within Computer Science that are deeply involved in the study of algorithmic fairness.

2.3. Participant recruitment & representation

Of relevance to our analysis is the systematic bias towards WEIRD participants found among Psychology studies (Henrich et al., 2010) and as recently also highlighted in studies published at CHI (Linxen et al., 2021). In an in-depth investigation across the behavioural sciences, Henrich et al. found that the majority of studies on human behaviour published in top journals are based on participant samples exclusively obtained from WEIRD societies (Henrich et al., 2010). This is in line with prior work from *e.g.* Arnett, who analysed the publications across six subdisciplines in Psychology between 2003–2007 and found that 68% of participants were from the USA, and 96% were recruited from Western industrialised countries (Arnett, 2008). Henrich et al. further found that members of these WEIRD societies are frequent outliers in domains such as fairness, moral reasoning, reasoning styles, and self-concepts (Henrich et al., 2010), all of which are highly relevant in shaping our perceptions of algorithmic fairness. With the above findings, Henrich et al. state that “WEIRD subjects may often be the worst population from which to make generalizations”, as this subgroup frequently acts as a significant outlier when compared to other global samples (Henrich et al., 2010, p. 79). Within CHI, Linxen et al. found that over the period 2016–2020, US participants account for 54.84% of all study participants and ‘Western’ participants (including the US) for 73.13%. 71% of participants were recruited in countries considered fully WEIRD (*i.e.*, Western and above the global median for Education, Industrialisation, Wealth, and Democracy).

In the context of fairness, relying on WEIRD participant samples can result in disparate outcomes. Blake et al. studied the development of two types of fairness decisions (disadvantageous inequity aversion and advantageous inequity aversion) across seven diverse societies (Blake et al., 2015). Their results show that while disadvantageous inequity aversion emerged around the same time across the studied populations, the development of advantageous inequity aversion (i.e., self receives more than peers) appeared only in three of the studied populations (Blake et al., 2015). Studies of adult populations have similarly found significant differences in perceptions of fairness and cooperation, with for example Herrmann et al. identifying a large variation in antisocial punishment between sixteen assessed participant pools (Herrmann et al., 2008). These examples from Psychology highlight how “WEIRD societies tend to be outliers on many measures of fairness and cooperation” (Blake et al., 2015, p. 258). In HCI, recent work by Nakao et al. points to the importance of considering culturally dependent perspectives in the development of fair AI systems—further stressing the importance of including studies from non-WEIRD societies (Nakao et al., 2022).

The relative ease and low cost with which researchers can recruit participants through crowdsourcing platforms, such as Amazon Mechanical Turk (MTurk) and Prolific, has led to a significant increase in the uptake of these platforms as a source of study participants (Irani and Silberman, 2013; Barbosa and Chen, 2019). While this paradigm has allowed researchers to recruit larger samples and enabled data labelling at a scale previously unattainable, research has also pointed to the potential bias this participant sample may introduce. Paolacci et al. found that workers on MTurk are generally younger, better educated, and have a more liberal worldview than the USA’s general population (Paolacci et al., 2010). According to data from MTurk Tracker for August 2020, roughly 75% of crowdworkers on this website are from the USA, 16% from India, and the remaining 10% originate from other countries (Difallah et al., 2018). While a young, highly educated, and more than average liberal US participant sample is perhaps unproblematic in many areas of HCI research (e.g., Fitts’s law experiments), findings from such a sample may differ significantly from other populations when discussing algorithmic fairness. Aitamurto & Chen point to the barriers that may limit participation in crowdsourcing efforts, such as a lack of internet access or a motivational barrier to participation (Aitamurto and Chen, 2017). We are, therefore, interested in assessing the reliance on crowdworkers for participant recruitment in the context of algorithmic fairness.

To overcome some of these biases introduced by crowdsourcing work, Barbosa and Chen proposed a framework that considers the demographics of the crowdworkers when distributing tasks (Barbosa and Chen, 2019). While the authors demonstrate that the framework was effective in reducing a potential bias of crowdworker demographics (country of origin, gender, and age), the authors also found that the use of their framework introduced a slight decrease in the accuracy of responses—possibly because the framework tasks did not recruit the most experienced workers to avoid country bias. Van Berkel et al. asked crowdworkers to discuss and indicate whether they believed that the use of a presented predictor variable was fair (van Berkel et al., 2019). Their results showed that discussions among a more diverse group (in terms of age, gender, race, and income) resulted in a closer alignment among group members with the overall majority than discussions in groups with lower diversity.

2.4. Reflections on HCI research

A systematic assessment of the literature within a community is a widely used technique to identify methodological practices. A recent example of such an analysis within HCI is a review of the ‘local standards for sample size’ by Caine (2016). Analysing all 465 papers of the CHI 2014 proceedings, Caine collected the sample size information of a total of 606 user studies, as well as participant demographics

and parameters related to the study designs. Analysing these results identified a wide spread in sample size (ranging from 1 to 960 000 participants), with 12 participants being the most commonly reported sample size (Caine, 2016). 71% of the study samples included college students. Following the identification of missing details, Caine recommends authors ensure that complete information about on the study’s sample size is provided, in part to facilitate future replication research (Caine, 2016). Liu et al. analysed the thematic evolution in the HCI field and identified underlying trends within the CHI community over two decades (Liu et al., 2014). Through a co-word analysis, Liu et al. identified *clusters* to assess the field’s interrelated concepts and intellectual structures. In addition to reporting on community’s growth, the authors also found that only 44.7% of the top keywords in the period between 1994–2003 were repeated in the papers published between 2004–2014, implying that the top research topics were replaced with new topics. For example, keywords such as ‘social networks’ and ‘crowdsourcing’ only emerged as research themes between 2004–2013 (Liu et al., 2014).

Several works have also reviewed the existing literature to learn about a specific topic within the HCI community. The work by Abdul et al. (2018) is most relevant for this paper. Through a topic modelling-based approach, in which the content of over 10 000 papers were analysed, the authors identified fading and emerging topics in the research field. Moreover, the scholars reveal research clusters and research communities and demonstrate how closely these entities relate to each other. Abdul et al.’s analysis shows that interdisciplinary research contributes the most towards progress in explainable systems (Abdul et al., 2018). Aiming to increase our understanding of the notion of ‘interaction’, Hornbæk et al. conducted an extensive literature review of 4604 papers published at CHI over the last 35 years (Hornbæk et al., 2019). Through a combined process of natural language processing and manual classification, the authors extracted n-grams and phrases that contain the word ‘interaction’ as mapping to relevant modifiers. The authors report an increased usage frequency of the term ‘interaction’ in papers in the later years of the conference as compared to the initial years. The authors also demonstrate more than 2000 modifications applied to the term ‘interaction’ throughout the conference proceedings in the last 35 years, indicating varying usage of the term ‘interaction’ (Hornbæk et al., 2019).

In this paper, we set out to explore how the topic of algorithmic fairness has been studied both in CHI and FAccT. Informed by the critiques brought forward in both other as well as our own scientific communities, and following an established practice in which conference proceedings are used to reflect on a community’s practice, we next present our systematic literature review.

3. Methodology

We conduct a comprehensive literature review on how researchers have studied user perceptions of algorithmic fairness. Our search focuses on papers published in the *Proceedings of the ACM Conference on Human Factors in Computing Systems* (hereafter, CHI) and the *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency* (hereafter, FAccT). The CHI conference is considered the premier conference in HCI, and provides an inclusive representation of the Human–Computer Interaction landscape (Hornbæk et al., 2019). Because of these characteristics, the CHI proceedings have been used previously to study practices and trends in the HCI community (Caine, 2016; Liu et al., 2014; Koeman, 2020; Pohl and Mottelson, 2019). The FAccT conference is a cross-disciplinary conference with a focus on the fairness, accountability, and transparency of socio-technical systems. A recent literature review highlights the most common topics within FAccT to focus on ‘Fairness and ML’, ‘Explainability’, and ‘Social Science, Ethics, and Accountability’ (Laufer et al., 2022), indicating a close alignment with our interest area of fairness perceptions.

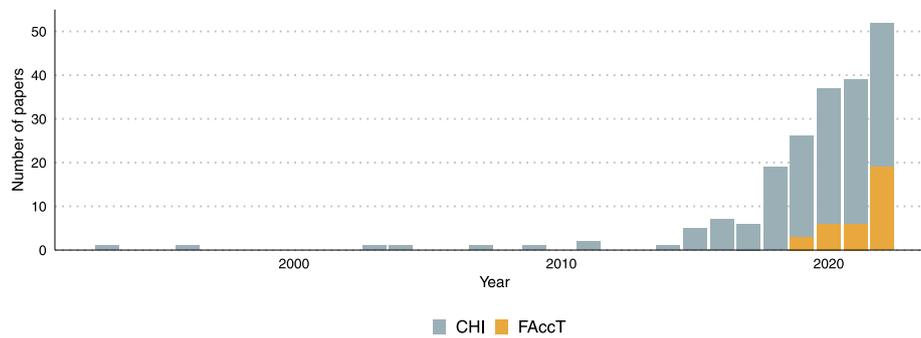


Fig. 1. Overview of frequency of papers on perceptions towards AI across CHI and FAccT.

We searched the *ACM Guide to Computing Literature* using an ‘AND’ search query combining terms surrounding fairness in an ‘inclusive or’; ‘fairness’, ‘equity’, ‘trust’, and ‘equality’, as well as terms related to algorithms in an ‘inclusive or’; ‘artificial intelligence’, ‘algorithm’. To be as exhaustive as possible in our search, we specified our search query to assess from the Digital Library both full-text and metadata (e.g., title, abstract, author keywords). We do not apply a date filter, as we are interested in mapping the evolution of this research area over time. The search was carried out in July 2022, covering the CHI proceedings from 1982 until 2022 and the FAccT proceedings from 2019 to 2022 (both years including).

Our search resulted in a total of 947 papers for CHI and 313 papers for FAccT. We individually analysed each paper to meet our inclusion criteria. First, we identified and excluded extended abstracts papers (e.g., workshop, tutorial). A total of 250 papers were excluded (225 for CHI, 25 for FAccT), bringing the remaining total to 1010 papers. Subsequently, we manually analysed each remaining document. Numerous papers contained references to AI conferences or journals, which led them to be incorrectly included in our search results. Papers which, for example, solely mentioned fairness-related aspects in ‘future works’ were excluded from our selection. and identified documents Further, papers which did not focus on human perceptions were excluded. This primarily concerned papers which presented a mathematical contribution without further evaluation with end-users. Literature reviews were also excluded from our analysis. Following these exclusion criteria, 166 CHI papers and 34 FAccT papers remained for further analysis.

4. Results

The 200 papers classified as relevant for our analysis were published between 1993 and 2022. As shown in Fig. 1, the number of papers published concerning algorithmic fairness has increased rapidly, with over 45% of the identified papers published in the last two years. Classifying the corpus of papers, we find that a majority of 136 papers, or 68% of the corpus, contain one user study, 49 papers contain two user studies, and 15 papers contain three or more user studies. For the subsequent analyses, we count each study individually. In other words, if a paper reports on two user studies, we consider these studies separately from each other. From the 200 total analysed papers, we identify a total of 283 studies at an average of 1.4 studies per paper. An overview of the characteristics of these 283 studies is presented in Appendix. Our analysis focuses on the study designs and methodological approaches, the studies’ participant samples, and the geographical location of both participants and study authors.

4.1. Study designs

We classified the *primary* research method for each study following the categorisation previously used by Koeman (2020). We describe these six categories below;

- **Lab studies.** Studies in which participants are invited to the lab and asked to complete a specific task.
- **Remote studies.** Studies in which participants are asked to complete a task remotely, most commonly involves crowdsourcing tasks and surveys.
- **Interview studies.** Studies in which participant data is collected through interviews and discussion.
- **Field studies.** Studies in which researchers study participants in the environment of interest, also known as *in situ* or in-the-wild studies.
- **Workshops.** Studies in which participants play an active role in generating study data, include, among others, co-design and expert evaluation studies.

In classifying the studies, we identified the primary methodology employed in the study, also following the aforementioned classification by Koeman (2020). If, for example, a field study was complemented with a final interview study, we mark it as a field study. We present an overview of the primary methodologies employed in each study in Table 1. We find that the most commonly used methods were remote studies (130 studies), interview studies (68 studies), and lab studies (41 studies). Fig. 2 highlights the distribution of research methods between CHI and FAccT studies in our corpus. While the analysed CHI papers included a substantial number of field studies (23 studies, 13.9% of analysed CHI papers), none of the analysed FAccT papers included a field study. The FAccT papers were also low on the number of lab studies (2, or 5.9%) compared to our CHI sample (39, or 23.5%).

4.1.1. Study duration

Next, we analyse the duration of the identified studies. In our classification, we uphold a minimum duration of one day (e.g., a 30-min interview study is classified as one day), and a study duration of three weeks is classified as $(3 \times 7 =) 21$ days. Studies in which duration differed between participants (e.g., participation between five to ten days (Wu and Munteanu, 2018)), we take the average study duration in our calculation. The mean study duration is 2.43 days. However, the median study duration of one day indicates that most studies are not longitudinal. The longest study in our sample has a duration of 180 days (Passi and Barocas, 2019). Table 1 provides an overview of the duration of studies across the various methodologies, with field studies predictably having the longest duration across our study sample.

4.1.2. Method of analysis

We identified the analyses performed in each study, classifying studies as either ‘qualitative’, ‘quantitative’, or ‘mixed method’ based on the primary form of analysis. Here, we focus specifically on how the results of a study are analysed, as opposed to the research method applied in the study (see Section 4.1). We categorised data analysis methods such as open coding of participant comments (e.g., Eslami et al. (2018)), interpretation of workshop results (e.g., Brown et al. (2019)), and the analysis of interview transcripts (e.g., Schneider et al. (2019))

Table 1
Study duration and sample size across the different study methodologies.

Methodology	Venue	No. studies	Mean duration (SD)	Median duration	Mean N (SD)	Median N
Lab study	CHI	39	1 (0)	1	30.6 (41.2)	20
	FAccT	2	1 (0)	1	11.0 (11.3)	11
	Total	41	1 (0)	1	29.6 (40.4)	19
Remote study	CHI	106	1.12 (1.26)	1	361 (623)	168
	FAccT	24	1 (0)	1	377 (619)	146
	Total	130	1.10 (1.14)	1	364 (620)	160
Interview study	CHI	54	1.02 (0.14)	1	26.9 (50.0)	17.5
	FAccT	14	13.80 (47.80)	1	19.1 (12.1)	14
	Total	68	1.01 (0.12)	1	25.4 (45.2)	16
Field study	CHI	23	19.4 (18.3)	14	39.6 (68.4)	18
	FAccT	0	-	-	-	-
	Total	23	19.4 (18.3)	14	39.6 (68.4)	18
Workshop	CHI	15	1.27 (1.03)	1	25.5 (22.6)	16
	FAccT	6	1.50 (1.22)	1	20.7 (10.1)	19
	Total	21	1.33 (1.06)	1	24.1 (19.7)	16
All	CHI	237	2.86 (7.86)	1	178.27 (448.66)	33
	FAccT	46	4.96 (26.39)	1	210 (482.47)	38
	Total	283	3.20 (12.80)	1	183.3 (453.48)	34

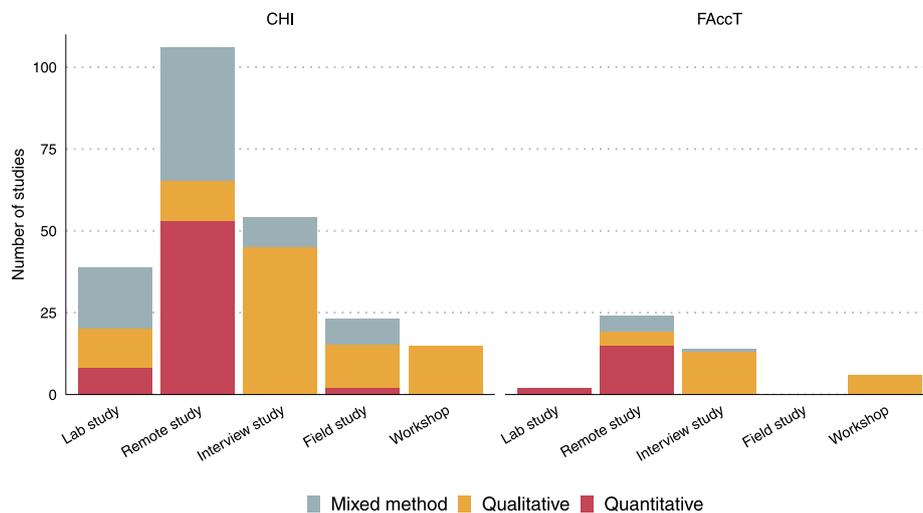


Fig. 2. Overview of analysis approaches across the identified methods.

as qualitative. Studies involving quantitative data analysis methods typically involved the use of inferential statistics across conditions in experimental designs (e.g., Yin et al. (2019)). Studies classified as mixed method combined a qualitative and quantitative analysis approach. For example, Verame et al. present a user study in which participants’ task performance is quantified and assessed, while participant strategies are identified through interviews (Verame et al., 2016). Our sample consists of 120 qualitative analyses, 80 quantitative analyses, and 83 mixed method analyses. Fig. 2 highlights the overlap between the study method and analysis technique as split between CHI and FAccT. Our results indicate that while remote studies typically rely on quantitative analysis, other research methods more often result in a qualitative or mixed method analysis. Unsurprisingly, we find extensive use of qualitative analysis methods for workshops and interview studies. We performed a Pearson’s Chi-squared test to examine the relation between publication venue and analysis method. The relationship between these variables was significant ($\chi^2(2, 283) = 7.18, p = 0.03$), with a posthoc test (using Bonferroni correction) highlighting a significantly lower number of mixed method studies in the FAccT conference ($p < 0.05$).

4.1.3. Compensation

In terms of participant compensation, we classified between ‘financial compensation’ (e.g., a set price for completion of the task, or

a raffle), vouchers, and other (e.g., study credits, physical gift). In a total of 127 studies (44.9%), financial compensation of participants was the most commonly used compensation technique. In 35 studies (12.4%) participants were rewarded with a voucher. Seven studies (2.5%) provided an alternative compensation method (Liebling et al., 2020; Thakkar et al., 2020; Hsu et al., 2021; Okolo et al., 2021; Setlur and Tory, 2022; Jakesch et al., 2022), and one study offered participants either a financial compensation or study credits (Völkel et al., 2020). For five studies the authors explicitly stated not to provide a reward for their participants (Jahanbakhsh et al., 2017; Liao et al., 2018; Hastings et al., 2020; Schußet al., 2021; Cheng et al., 2022). For a total of 108 studies (38.2%), the authors did not provide details on the compensation of participants. This represented 84 (35.4%) of CHI studies and 24 (47.8%) of FAccT studies.

4.2. Participant samples

Of the total number of 283 studies included in our sample, the average sample size is 183 (SD = 453.5) participants per study.² Given

² We were unable to identify the participant sample of one study (Passi and Barocas, 2019).

Table 2
Details on the four categories of participant roles as identified in the corpus.

Participant role	Venue	No. of studies	Mean <i>N</i>	Median <i>N</i>
Domain experts/Stakeholders	CHI	70	40.7	16
	FACcT	12	33.4	17
	Total	82	39.6	16
General public	CHI	78	43.6	24.5
	FACcT	5	182.0	47
	Total	83	51.9	25
Crowdworkers	CHI	71	497.0	237
	FACcT	15	513.0	392
	Total	86	500.0	276
HCI/ML practitioners	CHI	18	39.2	28
	FACcT	13	34.8	20
	Total	31	37.4	21

the inclusion of several studies with an extensive sample, the median sample size of 34 gives a better indication of the typical number of participants involved in the studies. The median sample size of the CHI and FACcT studies is largely comparable, at respectively 33 and 38 participants. Table 1 shows the mean and median sample size as split by the primary method of the study.

Following an initial read-through of the corpus of studies, the authors collectively decided on the following classification of participant roles in the studies:

- **Domain experts/Stakeholders.** The first category consists of participants specifically recruited due to their (professional) expertise or close personal involvement with the research goal. For example, studies focusing on a healthcare application in which healthcare workers are recruited belong to this category.
- **General public.** Studies in which participant recruitment is broad, and a selection is typically made based on participant availability as opposed to participant expertise. This includes recruitment of participants from University mailing lists as well as broad recruitment among social media users.
- **Crowdworkers.** Participants recruited through crowdsourcing websites such as Amazon Mechanical Turk or Prolific.
- **HCI/ML practitioners.** Studies in which recruited participants were HCI/ML practitioners or researchers.

Table 2 presents information on the total number of studies in which the aforementioned roles were involved, as well as the mean and median number of participants for each role. As seen from Table 2, studies involving crowdworkers as the participant sample had by far the largest sample size. Across all four categories of participants, considerable deviations in sample size occur, making the median sample size the most indicative of recruitment practices. We performed a Pearson’s Chi-squared test to assess the relationship between publication venue and participant role. The relation between these variables was significant ($\chi^2(3, 283) = 24.40, p < 0.01$). Posthoc tests (using Bonferroni correction) highlight a significantly lower number of FACcT studies involving members of the general public ($p = 0.02$) as compared to the analysed CHI papers, and a significantly higher number of FACcT studies involving HCI/ML practitioners as compared to the analysed CHI studies.

Further, we explored the distribution of participant roles across the different primary research methods as introduced in Section 4.1. Whereas the general public is responsible for most participation in lab studies, domain experts play an equally big role in field studies. Interview studies are split between participants recruited from the general public, domain experts, and HCI/ML practitioners. HCI and ML practitioners played a minor role in lab, remote, and workshop-based studies, but were not represented in field studies. Unsurprisingly, crowdworkers are found only in remote studies. We visualise this distribution of participant roles across research methods in Fig. 3.

4.2.1. Demographic variables

Next, we present our analysis of the demographic variables as reported by the studies’ authors. Specifically, we assess the reporting on participants’ gender, age, race, and education — four commonly used demographic variables.

Our results show that the aforementioned demographic variables are often not reported. While the majority of studies report information on their participants’ gender (64.0%), details regarding age (49.8%), education (24.4%), and race (12.4%) were often unreported. We conducted four separate Pearson’s Chi-squared tests to examine the relation between publication venue and the reporting of these four demographic variables. Our results show a significant relation between publication venue and the reporting of both gender ($\chi^2(1, 283) = 13.43, p < 0.001$) and age ($\chi^2(1, 283) = 9.21, p = 0.002$), with posthoc tests (using Bonferroni correction) highlighting a significantly higher number of reporting of gender and age demographics in CHI as compared to FACcT (respectively 68.8% vs. 39.1% for gender and 54.0% vs. 28.3% for age).

4.3. Geographical location

Finally, we assessed the geographic location of both study participants and authors within the identified study sample. First, we investigated the location from which participants were recruited. We note that recruitment location is not necessarily the participants’ nationality, but rather their geographical location at the time of the study. If the authors did not specify the location of their participants but noted that participants were recruited at the local university campus, we deduce participants’ country from the location of the authors’ institution. If a small minority of the population is from a second country (e.g., Uhde et al. report 3 Austrian participants and 48 German participants (Uhde et al., 2020)), we classified the sample population as per the overwhelming majority’s location. If there are one or more sizeable minorities, we report this as ‘multiple countries’. Participants were recruited primarily from the United States ($N = 95$), with Germany ($N = 11$) and the United Kingdom ($N = 9$) following by a large gap. A total of 30 studies recruited participants from multiple countries. For a total of 102 studies (36%) we could not identify the participants’ location.

We visualise the recruitment of participants in relation to the location of the first author’s institution in Fig. 4. To ensure the legibility of the figure, we group countries with fewer than five studies in the ‘Other countries’ category. The majority of studies recruit participants within the same country as the first author (75.1% of the studies for which information on participant location is available).

We next assess the degree of cross-country collaboration in the analysed corpus by analysing each paper’s author information. Here, we classify an author team as cross-country if at least one of the authors is from an institution residing in a different country. Of the 283 studies, 83 studies (29.3%) originate from a cross-country authorship team. A Pearson’s Chi-squared test to examine the relation between publication venue and cross-country authorship was significant ($\chi^2(1, 283) = 6.15, p = 0.01$), with a posthoc test (using Bonferroni correction) highlighting a significantly higher number of cross-country collaborative studies in the FACcT conference as compared to CHI ($p < 0.05$); 26.2% at CHI and 45.7% at FACcT. We further find that cross-country author teams were more likely to recruit their participant sample from multiple countries (13.2% for single-country author teams as compared to 25.0% for cross-country author teams). However, a Pearson’s Chi-squared test was not significant ($\chi^2(1, 181) = 2.94, p = 0.09$)—possibly due to the relatively low power due to few cross-country studies with participant recruited from multiple countries.

Finally, to assess whether the identified bias of participant location is due to differences in methods used, we evaluate the distribution of research methods between countries. We provide a visual overview of the distribution of primary research methods per country in Fig. 5.

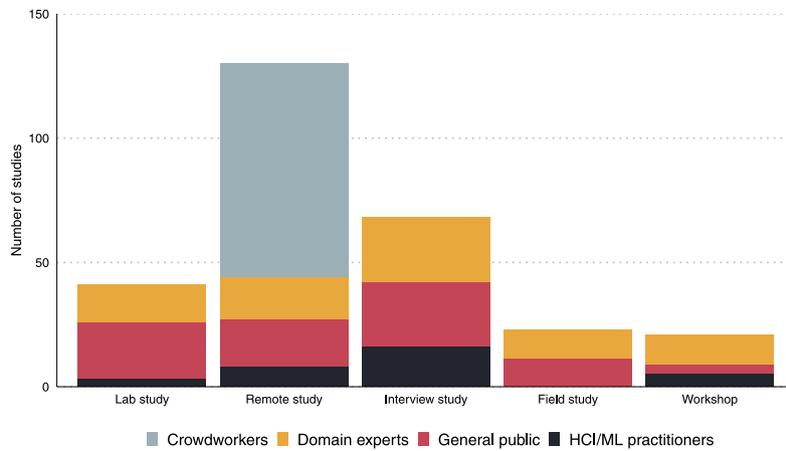


Fig. 3. Overview of participant roles across the identified methods.

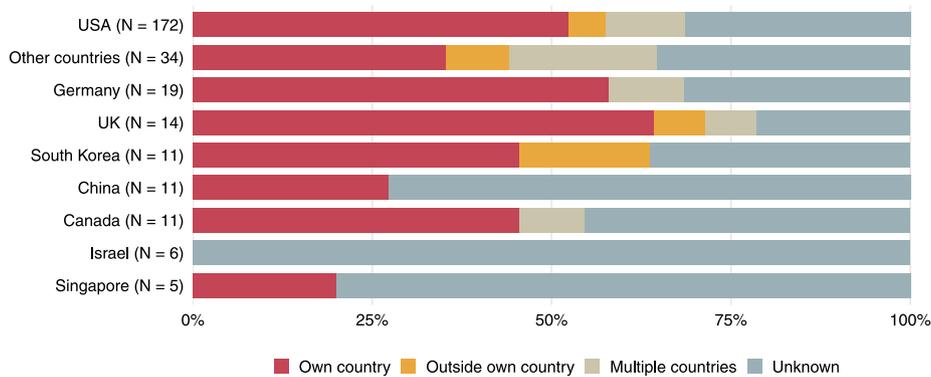


Fig. 4. Relationship between country of first author and country in which participants are recruited.

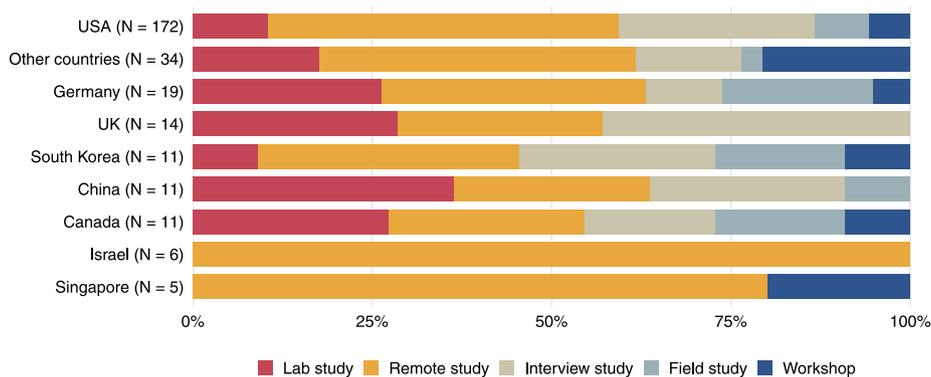


Fig. 5. Distribution of research methods employed, as grouped by the first author's country.

Again, we base our assessment on the location of the first author's institution. While the sample size is relatively small for some countries, this overview highlights the general diversity of methods applied by HCI researchers across the countries included in our sample. Furthermore, we show that the high number of US-only participant studies is not due to a more significant focus on lab studies in the US. On the contrary, authors from the US have a larger focus on remote studies and relatively

few lab studies compared to the samples from other countries. We note that the six remote studies attributed to Israel originate from one paper.

5. Discussion

Our analysis highlights the growing interest in the study of algorithmic fairness. As can be seen in Fig. 1, the number of papers on this topic within both CHI and FAccT is rapidly growing and can

be expected to be part of the research agenda for the foreseeable future. This increased interest is in line with earlier calls from the HCI community. For example, Abdul et al. state that “*the time is ripe for the HCI community to ensure that the powerful new autonomous systems have intelligible interfaces built-in*” (Abdul et al., 2018, p. 1), and Holstein et al. point to the “*critical opportunities for the ML and HCI research communities to play more active, collaborative roles in mitigating unfairness in real-world ML systems*” (Holstein et al., 2019, p. 12). Through an analysis of 200 papers on algorithmic fairness, we uncover common practices in how CHI and FAccT authors have studied perceptions of algorithmic fairness. Summarising the main takeaways across 283 studies, we find that most studies are cross-sectional (*i.e.*, measurement at a single point and short in duration). Regarding participant roles, we see that the CHI community recruits participants among domain experts/stakeholders, the general public, and crowdworkers roughly at an equal rate, but surprisingly has a minor focus on HCI/ML practitioners. The FAccT community, on the other hand, sees an under-representation of members of the general public in the analysed studies. We further find that US participants are disproportionately represented. Finally, we identified differences in the use of methods between countries, with US researchers more commonly carrying out remote studies than lab studies in comparison with other countries. Finally, we discuss the implications of our results in more detail below.

We note that critical information on the participant sample was frequently missing in the studies analysed. In 102 studies (36.0%), the location of participants was not explicitly stated and could not be reliably deduced from the authors’ descriptions. Knowing the location of participants is critical, as geographical differences can significantly impact perspectives on expected algorithmic behaviour. For example, Awad et al.’s ‘Moral Machine’ experiment, which collected data across 233 countries and territories, found distinct differences in participants’ moral viewpoints in relation to autonomous vehicles (Awad et al., 2018). Their results point to three distinct ‘moral clusters’ in their dataset, which align both with the geographical and cultural proximity of the participants’ countries. For 108 studies (38.2%), information on participants’ compensation was missing. The lack of information on participant compensation was particularly widespread among the analysed FAccT studies (47.8%, as compared to 35.4% at CHI). Previous work has studied the effect of compensation on participant data quality across a variety of studies (Stone et al., 1991; Musthag et al., 2011; Wiseman et al., 2017), as well as discussed the ethical considerations of sufficient participant compensation (Williamson, 2016). To interpret the results of a study correctly, as well as to support any future (replication) studies, details on participants’ location and compensation are critical. In addition, we find that demographic details of participant samples are often not reported, as assessed in terms of participants’ gender (64.0%), age (49.8%), education (24.4%), and race (12.4%). This is despite our lenient classification, in which we considered statements as “*a majority of participants [...]*” as sufficient information to assess a demographic variable reported.

How might the way in which researchers study algorithmic fairness affect algorithmic fairness outcomes? While it is outside of this paper’s scope to offer quantifiable evidence of differences between studied and unstudied/understudied populations, *e.g.* different geographical locations or population samples, our analysis reveals concrete gaps in our community’s study of algorithmic fairness. These gaps are most visible on a map in terms of geographical coverage, highlighted by the largely cross-sectional design of studies, and further amplified by a limited reporting on study sample details. Prior work analysing governmental AI policy points to distinct geographical differences in government considerations on dealing with fairness, ethics, and legislation of algorithmic technologies (Dexe and Franke, 2020; van Berkel et al., 2020). National governments point to pre-existing cultural conditions in shaping their AI strategy. For example, the Norwegian National Strategy for Artificial Intelligence states: “*Norwegian society is characterised by trust and respect for fundamental values such as human rights and privacy.*

This is something we perhaps take for granted in Norway, but leading the way in developing human-friendly and trustworthy artificial intelligence may prove a key advantage in today’s global competition (Norwegian Ministry of Local Government and Modernisation, 2019). Similarly, a recent report on the American AI initiative describes that; “*Continued American leadership in AI is of paramount importance to maintaining the economic and national security of the United States and to shaping the global evolution of AI in a manner consistent with our Nation’s values, policies, and priorities.*” (The White House - Office of Science and Technology Policy, 2020). In the following sections, we highlight how gaps in geographical coverage may limit the scope of algorithmic fairness research through a discussion of prior studies both within and outside of HCI.

5.1. Longitudinal and in situ versus cross-sectional and in vitro

The overwhelming majority of studies (91.9%) in our sample are cross-sectional studies, in which participants are interviewed, observed, or asked to perform a task for a short period (ranging from a few minutes to a couple of hours). These studies mostly take place *ex situ*. The HCI community has debated the benefit of *in situ* (also referred to as ‘in the wild’) studies in contrast to the additional effort and costs typically required in longitudinal *in situ* study design—typically focusing on usability studies (Kjeldskov et al., 2004; Nielsen et al., 2006; Rogers et al., 2007; Kjeldskov and Skov, 2014). Discussing the evolution of lab-based and field-based evaluations between 2004 and 2014, Kjeldskov & Skov conclude that “*mobile HCI research should move beyond focus on usability and usability evaluation, [...] we should embrace field studies that are truly wild and longitudinal in nature in order to fully experience and explore real world use.*” (Kjeldskov and Skov, 2014, p. 50). Based on our evaluation, such a methodological shift has not taken place within studies focused on algorithmic fairness.

Is the lack of longitudinal *in situ* studies in relation to algorithmic fairness critical? Prior work in Moral Psychology discusses the ‘fundamental attribution error’, which Harman describes as “*the error of ignoring situational factors and overconfidently assuming that distinctive behaviour or patterns of behaviour are due to an agent’s distinctive character traits*” (Harman, 1999, p. 1). Several studies have highlighted how ‘situational traits’ (or contextual factors) such as time pressure (Darley and Batson, 1973) or the presence of others (van IJzendoorn et al., 2010) impact participants’ (moral) behaviour. Therefore, capturing the effect of contextual factors is vital to understand how and when people’s consideration of algorithmic fairness changes (van Berkel et al., 2022). This requires a more significant emphasis on *in situ* studies within the domain of Human-AI interaction.

In interpreting our observations, it is worth contrasting our findings to earlier work assessing established research methods and practices. Caine presents an overview of the ‘local standards for sample size at CHI’ by analysing participant samples from the CHI 2014 conference (Caine, 2016), and Koeman examined methodological decisions of the CHI 2020 proceedings (Koeman, 2020). Caine’s and Koeman’s analyses focus on an individual proceeding of CHI, as opposed to a topic-specific analysis as presented in our paper. Their analyses, therefore, provide a good overview of research practices across a wide range of study goals. Koeman, in her analysis of CHI 2020 papers, found that “*Over 85% of studies studied participants for a day or less.*” (Koeman, 2020, p. 1). Our results indicate a similar distribution, with over 90% of studies lasting less than one day. Caine’s categorisation of methodologies is different from ours. Therefore, we classify Caine’s ‘diary study’, ‘experience sampling’, and ‘field study’ as ‘field study’ to allow a direct comparison with our categorisation (Caine, 2016). Using the aforementioned categorisation, Caine reports 6.4% of analysed studies to be field studies, similar to the 8.1% of studies categorised as field studies in our sample. We can therefore conclude that although studies on algorithmic fairness are rarely longitudinal and *in situ*, this pattern is largely in line with other research published at CHI. While critique on the lack of longitudinal studies in HCI is not new, see

e.g. Lazar et al.'s 'Research Methods in HCI' (Lazar et al., 2017), a direct consequence of the lack of longitudinal studies is that the effect of contextual factors on fairness perceptions, deemed as important in the aforementioned Psychology literature (Darley and Batson, 1973; Harman, 1999; van IJzendoorn et al., 2010), cannot be systematically captured and evaluated.

5.2. Geographical diversity

Of the 283 studies, we could deduce participants' location for less than two-thirds of the sample (180 studies). Of these, 97 study samples (53.9%) consist solely of US-based participants. What is the role of geographical location in the study of algorithmic fairness? Comparing the analysed studies suggests significant value in regional diversity in data collection. For example, Ahmed et al. analysed the use of an Indian ride-sharing application (Ola) (Ahmed et al., 2016), whereas Ma et al. and Lee et al. analysed the position of drivers in the USA ride-sharing economy (Uber/Lyft) (Ma et al., 2018; Lee et al., 2015). While all three papers investigate the driver's role in ride-hailing services, these works identify perspectives unique to their locale. Ahmed et al. point out that many Indian drivers did not have prior access to a smartphone or the internet, combine digital with traditional customer engagement methods, and experience uncertainty as to how the drivers' rating is calculated. The USA observations from Ma et al. and Lee et al. highlight distinctly different aspects, including a mismatch between the driver's desire for autonomy and the substantial control of the ride-hailing companies towards the drivers, as well as a critique on specific functionalities (e.g., UberPool, surge pricing). Similarities between the study samples include tight financial circumstances, which influence the drivers' behaviour and mistrust towards the algorithmic calculation of drivers' ratings. In another example found in our literature review, Liebling et al. studied the needs of three different populations (travellers based in the USA, migrant workers based in India, and immigrant populations in the USA) (Liebling et al., 2020). Their results highlight that current translation applications do not meet the needs of migrants, who they identified as the population with the highest translation needs, and point to the need for better support for low literacy users and broader dialect and accent support.

These examples highlight the value of geographically diverse participant samples, both at the level of an individual research paper and the broader research community. Some prior work in HCI and CSCW has made an effort to study technology use across cultures, often resulting in novel findings that cannot be identified by focusing on a culturally uniform participant sample (see e.g., Gao et al. (2017) and Baughan et al. (2021)). An illustrative example of this is the work by Gao et al. which studied affective grounding in the context of instant messaging through a study which involved both Chinese and US participants (Gao et al., 2017). Participants collaborated in pairs formed either by two people from the same country or two people from different countries. The findings from Gao et al.'s study highlight not only how cultural differences may make communication difficult (issues surrounding task approach or fluency), but also provide concrete recommendations for developing communication tools to promote collaboration between cultures (Gao et al., 2017).

A limited number of empirical studies have investigated regional differences concerning algorithmic fairness. Awad et al.'s 'Moral Machine experiment' is perhaps the most widely known example, in which participants indicated their preferred outcome between two scenarios depicting a (fatal) traffic accident involving an automated vehicle (Awad et al., 2018). Each set of scenarios shows a directly comparable setting. For example, a vehicle with three passengers that will hit a woman with a stroller as compared to a vehicle with three passengers that will hit a roadblock (and thereby avoid hitting the woman with a stroller). Their results highlight significant differences in the preferred behaviour of the vehicle between geographical and cultural clusters. Awad et al. identify a Western, Eastern, and Southern

cluster, with additional sub-clusters within the three main clusters (e.g., Commonwealth and Scandinavian countries) (Awad et al., 2018). Simultaneously, this study, as well as similar studies that provide a utilitarian moral reasoning perspective on autonomous car accidents, are critiqued due to their narrow scope in assessing ethical issues (JafariNaimi, 2018).

JafariNaimi describes that the presented case of the trolley problem (as used by Awad et al. (2018)) is that of 'quandary ethics', in which the parameters of a dilemma are predefined and fixed and choices are unambiguous (crash car into X or Y) (JafariNaimi, 2018). JafariNaimi argues that the case of autonomous driving accidents is not as simplistic as presented in these studies due to an inevitable level of uncertainty and fluidity in real-world contexts, the current framing of autonomous cars as part of an infrastructure designed around cars (not limited to physical infrastructure, but also e.g. legal infrastructure), and the lack of a long-term perspective on the effect of autonomous vehicles on our society (JafariNaimi, 2018). This critique by JafariNaimi provides a highly relevant and detailed outline of how algorithmic fairness extends beyond one specific decision, and instead is often part of a larger and more complex discussion. While quandary ethics are easily captured in logic decision making (*if-then*), it is critical for the larger HCI community to position the discussion of algorithmic fairness also in a larger context to consider the impacts of algorithmic decision making on all stakeholders.

5.2.1. Postcolonial computing

Our analysis highlights a lack of geographical diversity in participant samples, with 53.9% of identifiable participants located in the US. This is similar to Linxen et al.'s analysis of CHI participant samples between 2016 and 2020, in which 54.8% of participants are US-based (Linxen et al., 2021). The imbalance towards US participants also serves as a warning for the continued relevance of contemporary fairness research for a global audience. For example, our sample contains only two studies with participants from South America and none from Africa. Further, the studies poorly represent specific distinctive geographic and cultural clusters (e.g. Latin Europe, Nordic Europe, Eastern Europe, Middle East). This essentially excludes the perspective of individuals in these regions toward the development of fair algorithmic systems. Previous work has warned that the dominance of large USA technology companies could lead to 'digital colonialism' (Kwet, 2019), in which citizens in the Global South face "Big Tech corporations [that] control computer-mediated experiences, giving them direct power over political, economic, and cultural domains of life" (Kwet, 2019, p. 1). Within HCI, Irani et al. have described postcolonial studies as investigating "the historical transformation of conditions of cultural encounter" (Irani et al., 2010, p. 1311), highlighting how contemporary cultural encounters are shaped by "the history of global dynamics of power, wealth, economic strength, and political influence" (Irani et al., 2010, p. 1311).

Although the lack of geographical diversity of participants can be largely explained by the dominance of US authors (see Fig. 4), prior work highlights how such an in-balance can reinforce the notion of a dominant US viewpoint in which we perceive the US, and to a lesser degree, other Western viewpoints, as the default and position the perspective of other localities as 'Other' and in contrast to Western viewpoints. This notion is backed up by Kou et al.'s analysis of the description of study locations in CHI papers (Kou et al., 2018). Their analysis compared the mention of countries in the titles and text between studies conducted in Western and non-Western countries, and found that the papers describing studies conducted in non-Western countries were more likely to mention the country in the paper's title and text (Kou et al., 2018). Within the context of algorithmic fairness perceptions, Sambasivan et al. highlight how the current practice and study of fairness is heavily West-centric (Sambasivan et al., 2021a). The authors warn that "western AI fairness is becoming a universal ethical framework for AI" (Sambasivan et al., 2021a, p. 315), underlining the need for the recruitment of study participants outside the US.

While our analysis on participant locality was limited to a country-level analysis, we stress that studies conducted within one country, such as the US, can cover vastly different contexts, populations, and cultures. A recent example of this is found in work by Lee and Rich, who found that the mistrust among Black Americans in the US medical system also affects trust in medical AI systems (Lee and Rich, 2021). Lee and Rich stress the importance of assessing differences between social groups. Furthermore, Irani et al. in their CHI 2010 article on postcolonial computing, already highlighted the challenges that may arise in categorising and generalising culture across the existing scale of nation-states (Irani et al., 2010). For example, individuals with a migration history might uphold cultural values and norms from their current country and homeland. Similar to this limitation of our analysis at the country-level, the binary division into WEIRD and non-WEIRD societies is also highly restrictive as it overlooks between and within-country differences.

5.3. Contrasting CHI and FAccT

Our analysis includes a comparison between one of the largest HCI venues (CHI) and a growing community of researchers dealing explicitly with algorithmic fairness (FAccT). We next summarise the lessons learned for HCI researchers following this comparison. The most striking difference between analysed CHI and FAccT papers is the fact that almost double as many FAccT papers consist of a cross-country authorship team as compared to the assessed CHI papers. Our results show that cross-country authorship teams are twice as likely to have a participant sample from multiple countries. As such, HCI researchers should consider the implications of the overwhelming majority of studies being limited to individual countries. Cross-country collaborations could support the validation and further generalisation of empirical study results by recruiting participants across geographical and cultural borders. Such efforts could furthermore address the continued 'WEIRDness' of HCI research (Linxen et al., 2021).

We also identified several study-related aspects where the CHI community highlights higher diversity and compliance to methodological recommendations than the FAccT community. For example, using a mixed-method analysis approach was relatively rare in our sample of FAccT papers compared to the sample of CHI papers. This highlights a potential strength of HCI research, as a mixed method analysis can mitigate some of the limitations of an exclusively qualitative or quantitative analysis. Further, while both studied venues often lacked information regarding participant compensation and participant demographics, CHI papers more often provided this information than FAccT papers.

5.4. Recommendations for research on algorithmic fairness

Caine concluded, in analysing the CHI practices surrounding participant sample size, that “an understanding of community practice can complement existing methods of sample size determination” (Caine, 2016, p. 988). Similarly, understanding community practice can highlight shortcomings and areas for improvement in how a community conducts research. Surprisingly, our analysis highlights that participant details were frequently missing from the analysed studies. Details on the country where participants were recruited were missing in 36.0% of cases, and details on participant compensation in 38.2% of studies. These lacking details hamper researchers' ability to compare study results or replicate prior work. It is therefore critical for researchers to report details on recruitment strategy, including compensation and recruitment source (e.g., students, crowdworkers), as well as demographic factors of the participant sample (e.g., location, age distribution). Prior work has highlighted the potential impact of location on fairness perceptions (Awad et al., 2018). Compensation strategy was also found to influence study results (Musthag et al., 2011; Wiseman et al., 2017). As such, these parameters are essential to report in the context of studies on algorithmic fairness. We refer the interested reader to a

recent review of compensation practices in the HCI community (Pater et al., 2021).

Recommendation 1. Report details on the locality and compensation of study participants.

Basic demographic variables such as age and gender of participants were missing in respectively 50.2% and 36.0% of studies. The demographics of a study's population are valuable in identifying relations between demographic variables and the studied phenomena and understanding the limitations of the applicability of a study's findings. For example, the effect of gender on algorithmic fairness perceptions is still uncertain, with some studies identifying little to no impact (Grgić-Hlača et al., 2020) and others finding gender differences in beliefs about algorithmic fairness (Pierson, 2017). We do, however, not believe that all demographic variables are equally valuable to collect and report for each study. For example, the collection and reporting of race are highly complex (Hanna et al., 2020), and education classification varies significantly between countries. The relevance of given demographic variables must therefore be determined on an individual study basis.

Recommendation 2. Collect and report demographic variables that are of relevance to the study outcome.

As highlighted by Linxen et al. the fact that 73% of recent CHI study findings are based on Western participant samples is primarily due to the predominance of Western authorship (Linxen et al., 2021). As further validated across our study sample, most HCI studies recruit participant samples locally (Fig. 4). To change this practice of mostly local recruitment would raise a myriad of ethical (see e.g. the discussion on the history of anthropology (Brightman and Grotti, 2020)), financial (e.g., demands by funding agencies), ecological (e.g., international travel), and practical concerns, and would set to disconnect researchers from their local communities. We, therefore, do not advocate for researchers to drop their current practice of local participant recruitment. Instead, reflecting on the obtained results concerning the studied participant population can enhance study results, including the study's applicability to other settings.

While it is not uncommon for papers' limitation sections to highlight that results are unlikely to extrapolate to other user groups, a deeper level of engagement with the specifics of the study sample can result in a richer understanding of algorithmic fairness perceptions. An example of this can be found in work by Holstein et al. which highlights that local US regulations prohibit access to sensitive demographic data often proposed in fair ML auditing methods (Holstein et al., 2019). Similarly, Sambasivan et al.'s cross-country analysis finds that the price sensitivity and recentness of AI technology in new markets may increase the predominance of issues with data quality (Sambasivan et al., 2021b). The fact that one-third of the identified papers do not indicate the geographical location of their participants highlights that there is room for improvement in this area.

Recommendation 3. Reflect on the implications of the participant sample on study findings related to algorithmic fairness. Avoid assumptions of US/Western countries as the 'default' context.

For the large majority of remote studies, participants were recruited from crowdsourcing platforms (Fig. 3). Whereas researchers from the US rely more heavily on remote studies than their peers (Fig. 5), this has not resulted in significant recruitment of participants from outside of the US or across multiple countries (see Fig. 4). In multiple studies we came across (US) authors arguing for a geographical limitation of their crowdsourcing sample to the US only to ensure sufficient English language understanding. Given the widespread understanding of the English language, including both native and non-native speakers, alternative recruitment filters to assess participants' language skills can increase the geographical diversity of remote study participants.

Recommendation 4. Motivate or eliminate country-level restrictions in crowdsourced participant recruitment to obtain a more globally diverse participant sample.

Our analysis shows that only 23 of the analysed papers collect data from more than one country. As a consequence of this and the generally low numbers of replication studies (Hornbæk et al., 2014; Echter and Häußler, 2018), we are mostly unaware of whether the identified findings hold in other countries. This is often a highly burdensome or even impossible task for individual research groups, with the aforementioned work from Awad et al. as a rare example of cross-country insight on algorithmic fairness perceptions (Awad et al., 2018). As such, there is a need for HCI researchers to develop collaborative methods and practices which enable cross-country comparisons of algorithmic fairness perceptions. Our results indicate that studies with a cross-country author team more often included participants from multiple countries (25.0%) than author teams originating from a single country (13.2%).

A successful example of large-scale collaboration within Psychology is the Psychological Science Accelerator (PSA). PSA is a 'distributed laboratory network' with over 500 globally distributed psychological science laboratories working in collaboration, with one of the aims being to recruit culturally and geographically diverse study samples (Moshontz et al., 2018). No efforts of this scale currently exist within HCI. While required coordination efforts are colossal, study results could reach a more globally diverse participant sample than studies currently found in HCI. Such a process is, naturally, a long-term community effort of which the specifics go beyond the scope of this paper. The possibility of such a shared research effort could be discussed at a future workshop or similar venue. Within the relatively novel research domain of algorithmic fairness, such a study could, for example, assess the degree to which earlier attempts at quantifying fairness perceptions replicate.

Recommendation 5. Promote cross-country collaboration in order to reach more geographically diverse study samples.

Finally, our sample highlights a lack of longitudinal studies, which inhibits our ability to assess the effect of contextual factors on algorithmic fairness perceptions (van Berkel et al., 2022). Within Psychology, contextual factors of e.g. time and social company have been shown to affect our moral judgement. For example, Kouchaki & Smith found that participants were less likely to lie or cheat in completing prescribed tasks in the morning than in the afternoon (Kouchaki and Smith, 2014). Yudkin et al. studied the effect of social presence on the importance afforded to various moral values through a longitudinal experience-sampling study (Yudkin et al., 2019). Participant labelled data on their current social company highlighted that people rate moral values as more important when in the presence of others with whom they have a close connection. While the current lack of longitudinal studies makes it impossible to assess the impact of these findings on algorithmic fairness perceptions, they are a worthwhile and relevant avenue for further exploration. An example of a field study in our sample is the work by Wu and Munteanu (2018). Their paper presents a study on user acceptance of a wearable device for fall risk assessments. The authors conclude that offering participants contextual data on their fall risk estimation increased end-user acceptance and perceived usefulness of the technology, highlighting the impact of context on user perceptions.

Recommendation 6: Followup cross-sectional lab studies with longitudinal field studies to investigate the impact of context on participants' fairness perceptions.

5.5. Limitations

We recognise a number of limitations in our work. First, as in any literature review, the identified papers are the direct result of our search criteria. Our search terms, 'fairness', 'equity', 'trust', or 'equality' – AND – 'artificial intelligence' or 'algorithm', were selected to include a wide range of papers, resulting in a total of 1260 papers from which we manually identified 200 relevant papers. Second, we limited our review to papers published at the CHI and FAccT conferences. While

the inclusion of more specialised and distinct conferences, such as e.g. AutoUI or CSCW, would allow for a broader representation of HCI research, this increases the risk of biasing our results towards a sub-domain of HCI. Third, and in line with our scope of publication venues, we analysed methodological aspects without considering the domain where the studies took place (e.g., healthcare, transportation). An analysis of (fairness) perceptions of AI in specific domains would be a valuable avenue for future work. Finally, we stress that the location information reported in a paper does not represent the cultural background of the study's participants or its authors.

To support extensions and replications of our work, we provide a .CSV file containing the extracted parameters from the 200 papers in our corpus: <https://osf.io/7dfz5/>.

6. Conclusion

This paper provides an analysis of the methodological considerations of 213 studies on the topic of algorithmic fairness. Given the growing interest in this domain, as seen both within the CHI and FAccT communities (see Fig. 1), a better understanding of how we study this critical area is highly important. Such an understanding provides newcomers to the field an opportunity to better grasp the different perspectives on this interdisciplinary domain, but most importantly allows for inner reflection within our community. Our analysis revealed several 'blind spots' in contemporary work on algorithmic fairness, including a lack of geographical diversity, few longitudinal studies, and the under-reporting of essential study information. While the recruitment of a more geographically diverse study sample, as well as an increased focus on longitudinal investigations, will undoubtedly require a significant effort, such work is necessary to strengthen our community's contribution to the global deployment of fair and human-centred algorithmic systems.

CRedit authorship contribution statement

Niels van Berkel: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Zhanna Sarsenbayeva:** Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Jorge Goncalves:** Formal analysis, Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The analysed corpus of studies is made publicly available and can be found at <https://osf.io/7dfz5/>.

Acknowledgement

Niels van Berkel received support from the Carlsberg Foundation, Denmark, Young Researcher Fellowship project 'Algorithmic Explainability for Everyday Citizens'.

Appendix

See Table 3.

Table 3
Overview of the study's characteristics as collected in the review.

Conf.	Reference	Year	Lab study	Remote study	Interview study	Field study	Workshop	Duration	Qualitative	Quantitative	Mixed method	N	Domain experts	General public	Crowdworkers	HCI/ML pract.	Participant country	Gender	Race	Age	Education	Author country	Author crosscountry		
CHI	Maulsby et al. (1993)	1993	•					1	•			8		•									CAN	•	
CHI	King and Ohya (1996)	1996	•					1		•		18		•					•					USA	•
CHI	Cosley et al. (2003)	2003		•				1		•		536		•										USA	
CHI	Lee et al. (2004)	2004	•					1	•	•		40		•			USA	•		•				USA	
CHI	Lee et al. (2004)	2004	•					1	•	•		20		•			USA	•		•				USA	
CHI	Tullio et al. (2007)	2007				•		42			•	13	•				USA							USA	
CHI	Lim et al. (2009)	2009		•				1		•		53							•	•	•			USA	
CHI	Lim et al. (2009)	2009		•				1		•		158			•				•	•	•			USA	
CHI	Bateman et al. (2011)	2011	•					1		•		24		•			CAN	•		•	•			CAN	
CHI	Yamamoto and Tanaka (2011)	2011		•				1		•		960			•				•	•	•			JPN	
CHI	Yamamoto and Tanaka (2011)	2011		•				1		•		10		•			JPN							JPN	
CHI	Solomon (2014)	2014		•				1		•		99			•				•	•	•			USA	
CHI	Eslami et al. (2015)	2015				•		1			•	40		•			USA	•	•	•				USA	
CHI	Lee et al. (2015)	2015		•				1	•			33	•				USA	•		•				USA	
CHI	Loepp et al. (2015)	2015	•					1		•		33		•					•	•	•			DEU	
CHI	Rader and Gray (2015)	2015		•				1	•			464			•		USA	•	•	•				USA	
CHI	Warsaw et al. (2015)	2015	•					1		•		18		•			USA	•		•				USA	•
CHI	Ahmed et al. (2016)	2016				•		14	•			66	•				IND	•		•	•			USA	•
CHI	Ashktorab and Vitak (2016)	2016					•	5	•			21		•			USA	•		•	•			USA	•
CHI	Depping et al. (2016)	2016	•					1		•		42		•										CAN	
CHI	Kizilcec (2016)	2016		•				1		•		103	•						•	•	•			USA	
CHI	Luger and Sellen (2016)	2016				•		1	•			14		•			GBR	•		•	•			GBR	
CHI	Verame et al. (2016)	2016	•					1		•		60		•			GBR	•		•	•			GBR	
CHI	Yang et al. (2016)	2016				•		13	•			24	•				USA	•		•	•			USA	
CHI	Colley et al. (2017)	2017				•		1	•			375	•				Mult	•		•				FIN	•
CHI	Jahanbakhsh et al. (2017)	2017		•				1	•			277	•				USA	•		•	•			USA	
CHI	Jahanbakhsh et al. (2017)	2017				•		1	•			13	•				USA	•		•	•			USA	
CHI	Lee et al. (2017)	2017		•				1	•			31	•				USA	•	•	•				USA	
CHI	MacLeod et al. (2017)	2017		•				1		•		6	•						•	•	•			USA	
CHI	MacLeod et al. (2017)	2017		•				1		•		100	•				USA	•		•	•			USA	
CHI	Moritz et al. (2017)	2017	•					1	•			5				•								USA	
CHI	Moritz et al. (2017)	2017		•				1	•			3				•								USA	
CHI	Ur et al. (2017)	2017		•				1		•		4509			•		USA	•		•	•			USA	
CHI	Alvarado and Waern (2018)	2018				•		1	•			11		•										CRI	•
CHI	Alvarado and Waern (2018)	2018				•		1	•			8		•										CRI	•
CHI	Binns et al. (2018)	2018	•					1		•		19		•			GBR	•		•	•			GBR	
CHI	Binns et al. (2018)	2018		•				1		•		325		•			GBR	•		•	•			GBR	
CHI	Binns et al. (2018)	2018		•				1		•		65		•			GBR	•		•	•			GBR	
CHI	Dolin et al. (2018)	2018		•				1		•		306		•					•	•	•			USA	
CHI	Dolin et al. (2018)	2018		•				1		•		237		•					•	•	•			USA	
CHI	Eslami et al. (2018)	2018	•					1	•			32		•			USA	•	•	•				USA	
CHI	Flintham et al. (2018)	2018		•				1	•			309		•			GBR	•		•	•			GBR	
CHI	Flintham et al. (2018)	2018	•					1	•			9		•			GBR	•		•	•			GBR	
CHI	Hamidi et al. (2018)	2018				•		1	•			13	•						•	•	•			USA	
CHI	Hu et al. (2018)	2018		•				1	•	•		604			•									USA	
CHI	Le Bras et al. (2018)	2018				•		1	•	•		10		•			GBR	•		•	•			GBR	
CHI	Liao et al. (2018)	2018				•		38.5		•		337	•				USA	•		•	•			USA	•
CHI	Marathe and Toyama (2018)	2018				•		1	•			5	•				USA	•		•	•			USA	
CHI	Marathe and Toyama (2018)	2018				•		1	•			10	•				USA	•		•	•			USA	
CHI	Rader et al. (2018)	2018		•				1		•		681			•		USA	•		•	•			USA	
CHI	Skirpan et al. (2018)	2018		•				1		•		175			•				•	•	•			USA	
CHI	Swearngin et al. (2018)	2018	•					1		•		16	•						•	•	•			USA	
CHI	Vaccaro et al. (2018)	2018		•				1		•		32		•			USA	•	•	•				USA	
CHI	Vaccaro et al. (2018)	2018		•				1		•		106			•		USA	•		•	•			USA	
CHI	Veale et al. (2018)	2018				•		1	•			27			•		Mult	•		•	•			GBR	
CHI	Verma and Dombrowski (2018)	2018		•				1	•			16	•				USA	•		•	•			USA	
CHI	Woodruff et al. (2018)	2018				•		1	•			44	•				USA	•		•	•			USA	
CHI	Wu and Munteanu (2018)	2018				•		1	•			5	•				CAN	•		•	•			CAN	
CHI	Wu and Munteanu (2018)	2018				•		7.5	•			4	•				CAN	•		•	•			CAN	
CHI	Xu et al. (2018)	2018	•					1		•		18	•						•	•	•			CHN	•
CHI	Amershi et al. (2019)	2019	•					1	•			49			•		Mult	•		•	•			USA	
CHI	Amershi et al. (2019)	2019		•				1	•			11			•		Mult	•		•	•			USA	
CHI	Ashktorab et al. (2019)	2019		•				1		•		203			•		Mult	•		•	•			USA	•
CHI	Barbosa and Chen (2019)	2019		•				1		•		1919			•		Mult	•		•	•			USA	
CHI	Braun et al. (2019)	2019				•		1	•			55		•			DEU	•		•	•			DEU	•
CHI	Brown et al. (2019)	2019				•		1	•			83	•				USA	•		•	•			NZL	•
CHI	Cheng et al. (2019)	2019		•				1		•		199			•		USA	•		•	•			USA	

(continued on next page)

Table 3 (continued).

Conf.	Reference	Year	Lab study	Remote study	Interview study	Field study	Workshop	Duration	Qualitative	Quantitative	Mixed method	N	Domain experts	General public	Crowdworkers	HCI/ML pract.	Participant country	Gender	Race	Age	Education	Author country	Author crosscountry	
CHI	Ding et al. (2019)	2019			•	•		21	•			10		•			CHN	•		•		CHN	•	
CHI	Eslami et al. (2019)	2019			•			1	•			15		•			USA	•	•	•	•		USA	
CHI	Holstein et al. (2019)	2019			•			1		•		35				•							USA	
CHI	Holstein et al. (2019)	2019		•				1			•	267				•							USA	
CHI	Jakesch et al. (2019)	2019		•				1		•		389			•		USA	•		•			USA	
CHI	Jakesch et al. (2019)	2019		•				1		•		196			•		USA	•		•			USA	
CHI	Jakesch et al. (2019)	2019		•				1		•		208			•		USA	•		•			USA	
CHI	Kittley-Davies et al. (2019)	2019	•					1			•	40		•			GBR	•		•			GBR	
CHI	Koch et al. (2019)	2019	•					1			•	16	•				FIN	•		•	•		FIN	
CHI	Kocielnik et al. (2019)	2019		•				1		•		116			•		USA						USA	
CHI	Kocielnik et al. (2019)	2019		•				1		•		325			•		USA						USA	
CHI	Kuhlman et al. (2019)	2019		•				1		•		144			•								USA	
CHI	Kunkel et al. (2019)	2019				•		14			•	93		•						•	•		DEU	
CHI	McCormack et al. (2019)	2019	•					1		•		7	•							•	•		AUS	•
CHI	Roy et al. (2019)	2019		•				1		•		781			•					•	•		CAN	
CHI	Schneider et al. (2019)	2019				•		28	•			9	•				DEU	•		•	•		DEU	
CHI	Sun et al. (2019)	2019			•			1	•			32	•				USA	•			•		USA	
CHI	Sundar and Kim (2019)	2019		•				1		•		157			•		USA	•	•	•	•		USA	
CHI	Wang et al. (2019b)	2019				•		1	•			14	•				SGP	•					SGP	•
CHI	Wang et al. (2019a)	2019			•			1	•			6	•										CHN	•
CHI	Wang et al. (2019a)	2019	•					1	•			2	•										CHN	•
CHI	Wang et al. (2019a)	2019	•					1		•		13	•							•	•	•	CHN	•
CHI	Yin et al. (2019)	2019		•				1		•		1994			•		USA						USA	
CHI	Yin et al. (2019)	2019		•				1		•		757			•		USA						USA	
CHI	Yin et al. (2019)	2019		•				1		•		1042			•		USA						USA	
FAccT	Passi and Barocas (2019)	2019			•			180	•			?			•		USA						USA	
FAccT	Green and Chen (2019)	2019		•				1		•		554			•		USA	•	•	•	•		USA	
FAccT	Lai and Tan (2019)	2019		•				1		•		480			•		USA						USA	
CHI	Andalibi and Buss (2020)	2020			•			1	•			13		•			USA	•	•	•			USA	
CHI	Beede et al. (2020)	2020				•		3	•			13	•				THA	•					USA	•
CHI	Chin et al. (2020)	2020		•				1		•		37		•						•	•		KOR	
CHI	Chin et al. (2020)	2020	•					1		•		94								•	•		KOR	
CHI	Cryan et al. (2020)	2020		•				1		•		1097			•		USA	•					USA	
CHI	Cryan et al. (2020)	2020		•				1		•		203			•		USA	•					USA	
CHI	Diana et al. (2020)	2020		•				1		•		60			•				•	•			USA	
CHI	Dillen et al. (2020)	2020			•			1		•		20		•			CAN	•		•			CAN	
CHI	Fan and Zhang (2020)	2020		•				1		•		82			•					•	•		USA	
CHI	Geeng et al. (2020)	2020			•			1	•			25		•			USA						USA	
CHI	Gero et al. (2020)	2020	•					1	•			11	•				USA	•					USA	
CHI	Gero et al. (2020)	2020		•				1		•		89			•					•	•		USA	
CHI	Hastings et al. (2020)	2020		•				1		•		289	•				USA	•			•		USA	•
CHI	Hastings et al. (2020)	2020			•			1	•			18	•				USA				•		USA	•
CHI	Hong et al. (2020)	2020		•				1		•		100			•		USA	•		•			USA	
CHI	Kim et al. (2020)	2020		•				1		•		89		•			KOR						KOR	•
CHI	Kim et al. (2020)	2020				•		21		•		32		•			KOR						KOR	•
CHI	Kontogiorgos et al. (2020)	2020	•					1		•		44		•			SWE	•		•			SWE	
CHI	Lai et al. (2020)	2020	•					1	•			16		•			USA	•		•			USA	
CHI	Lai et al. (2020)	2020		•				1		•		480			•		USA	•					USA	
CHI	Lai et al. (2020)	2020		•				1		•		480			•		USA	•					USA	
CHI	Lai et al. (2020)	2020		•				1		•		480			•		USA	•		•			USA	
CHI	Liang et al. (2020)	2020			•			7	•			18		•						•	•		USA	•
CHI	Liang et al. (2020)	2020			•			1	•			2	•										USA	•
CHI	Liao et al. (2020)	2020		•				1	•			20			•		USA	•					USA	
CHI	Liebling et al. (2020)	2020		•				1		•		3105			•		USA						USA	
CHI	Liebling et al. (2020)	2020			•			1	•			16		•			IND						USA	
CHI	Liebling et al. (2020)	2020			•			1	•			9		•			USA	•		•			USA	
CHI	Louie et al. (2020)	2020	•					1		•		21	•				USA	•		•			USA	
CHI	Madaio et al. (2020)	2020				•		1	•			48			•					•			USA	•
CHI	Mallari et al. (2020)	2020		•				1		•		1600			•								USA	
CHI	Mallari et al. (2020)	2020		•				1		•		1600			•								USA	
CHI	Oh et al. (2020)	2020	•					1		•		30		•			KOR	•		•			USA	•
CHI	Schaeckermann et al. (2020)	2020		•				1		•		12	•				Mult						CAN	
CHI	Smith et al. (2020)	2020		•				1	•			16	•										USA	
CHI	Smith-Renner et al. (2020)	2020		•				1		•		180			•		USA	•		•			USA	
CHI	Smith-Renner et al. (2020)	2020		•				1		•		180			•		USA	•		•			USA	
CHI	Sun et al. (2020)	2020			•			1	•			4	•										CHN	
CHI	Sun et al. (2020)	2020	•					1	•			2	•										CHN	
CHI	Sun et al. (2020)	2020			•			1	•			6	•										CHN	

(continued on next page)

Table 3 (continued).

Conf.	Reference	Year	Lab study	Remote study	Interview study	Field study	Workshop	Duration	Qualitative	Quantitative	Mixed method	N	Domain experts	General public	Crowdworkers	HCI/ML pract.	Participant country	Gender	Race	Age	Education	Author country	Author crosscountry
CHI	Thakkar et al. (2020)	2020						1				38					IND					IND	
CHI	Uhde et al. (2020)	2020						1				3					DEU					DEU	
CHI	Uhde et al. (2020)	2020						1				51					DEU					DEU	
CHI	Völkel et al. (2020)	2020						1				21					DEU					DEU	
CHI	Wang et al. (2020)	2020						1				579					USA					USA	
CHI	Xie et al. (2020)	2020						1				77					USA					USA	
CHI	Xie et al. (2020)	2020						1				3					USA					USA	
CHI	Xie et al. (2020)	2020						1				6					USA					USA	
CHI	Yan et al. (2020)	2020						1				30										USA	
FAccT	Marcinkowski et al. (2020)	2020						1				304					DEU					DEU	
FAccT	Harrison et al. (2020)	2020						1				502					USA					USA	
FAccT	Noriega-Campero et al. (2020)	2020						1				14										USA	
FAccT	Zhang et al. (2020)	2020						1				72										USA	
FAccT	Zhang et al. (2020)	2020						1				9										USA	
FAccT	Mustafaraj et al. (2020)	2020						1				392					USA					USA	
FAccT	Lucic et al. (2020)	2020						1				75					NLD					NLD	
CHI	Anik and Bunt (2021)	2021						1				17					CAN					CAN	
CHI	Anik and Bunt (2021)	2021						1				27										CAN	
CHI	Bae Brandtzæg et al. (2021)	2021						14				16					NOR					NOR	
CHI	Bennett et al. (2021)	2021						1				25										USA	
CHI	van Berkel et al. (2021)	2021						1				75					USA					DEN	
CHI	Cheng et al. (2021)	2021						1				12					USA					USA	
CHI	Crisan and Fiore-Gartland (2021)	2021						1				29										USA	
CHI	Ehsan et al. (2021)	2021						1				29										USA	
CHI	Gilad et al. (2021)	2021						1				362										ISR	
CHI	Gilad et al. (2021)	2021						1				375										ISR	
CHI	Gilad et al. (2021)	2021						1				323										ISR	
CHI	Gilad et al. (2021)	2021						1				361										ISR	
CHI	Gilad et al. (2021)	2021						1				93										ISR	
CHI	Gilad et al. (2021)	2021						1				79										ISR	
CHI	Hsu et al. (2021)	2021						1				49					USA					USA	
CHI	Hsu et al. (2021)	2021						1				22					USA					USA	
CHI	Jacobs et al. (2021)	2021						1				10										USA	
CHI	Jacobs et al. (2021)	2021						1				8										USA	
CHI	Jiang et al. (2021)	2021						1				36					AUS					AUS	
CHI	Lee and Rich (2021)	2021						1				187					USA					USA	
CHI	Lee and Rich (2021)	2021						1				21					USA					USA	
CHI	Levy et al. (2021)	2021						1				18					USA					USA	
CHI	Liao and Sundar (2021)	2021						1				293										USA	
CHI	Lima et al. (2021)	2021						1				200					USA					KOR	
CHI	Lima et al. (2021)	2021						1				194					USA					KOR	
CHI	Mendez et al. (2021)	2021						1				12					ECU					ECU	
CHI	Mendez et al. (2021)	2021						1				91					ECU					ECU	
CHI	Okolo et al. (2021)	2021						1				21					IND					USA	
CHI	Park et al. (2021)	2021						1				21										KOR	
CHI	Rahman et al. (2021)	2021						1				256					BGD					BGD	
CHI	Rietz and Maedche (2021)	2021						1				6					DEU					DEU	
CHI	Rietz and Maedche (2021)	2021						1				11					DEU					DEU	
CHI	Robertson et al. (2021a)	2021						1				13					USA					USA	
CHI	Robertson et al. (2021b)	2021						1				15					Mult					USA	
CHI	Robertson et al. (2021b)	2021						1				259					Mult					USA	
CHI	Sambasivan et al. (2021b)	2021						1				53					Mult					USA	
CHI	Samrose et al. (2021)	2021						1				120										USA	
CHI	Samrose et al. (2021)	2021						28				49										USA	
CHI	Schneider et al. (2021)	2021						1				40					DEU					DEU	
CHI	Schufset al. (2021)	2021						1				11					DEU					DEU	
CHI	Tahir et al. (2021)	2021						1				95										SAU	
CHI	Tahir et al. (2021)	2021						1				46										SAU	
CHI	Tsai et al. (2021)	2021						1				25					USA					USA	
CHI	Tsai et al. (2021)	2021						1				20					USA					USA	
CHI	Wang et al. (2021a)	2021						1				30					Mult					USA	
CHI	Wang et al. (2021b)	2021						7				22					CHN					USA	
CHI	Widder et al. (2021)	2021						70				17					USA					USA	
CHI	You et al. (2021)	2021						1				30					CHN					USA	
CHI	Zehrung et al. (2021)	2021						1				114										USA	
FAccT	Celis et al. (2021)	2021						1				76					USA					USA	
FAccT	Shen et al. (2021)	2021						1				56					USA					USA	
FAccT	Miceli et al. (2021)	2021						1				15					Mult					GER	
FAccT	Miceli et al. (2021)	2021						1				14					Mult					GER	
FAccT	Andrus et al. (2021)	2021						1				38					Mult					USA	

(continued on next page)

Table 3 (continued).

Conf.	Reference	Year	Lab study	Remote study	Interview study	Field study	Workshop	Duration	Qualitative	Quantitative	Mixed method	N	Domain experts	General public	Crowdworkers	HCI/ML pract.	Participant country	Gender	Race	Age	Education	Author country	Author crosscountry	
FAccT	Kasinidou et al. (2021)	2021		*				1			*	99					Mult	*	*	*	*	CYP		
FAccT	Jesus et al. (2021)	2021	*					1		*		3	*										PRT	
CHI	Kapania et al. (2022)	2022			*			1	*			32		*			IND	*	*	*	*	IND	*	
CHI	Kapania et al. (2022)	2022		*				1			*	459			*		IND	*	*	*	*	IND	*	
CHI	Tolmeijer et al. (2022)	2022		*				1		*		428			*		Mult	*	*	*	*	CHE		
CHI	Park et al. (2022)	2022				*		1	*			50	*										KOR	*
CHI	Kim et al. (2022a)	2022				*		25	*			36	*				KOR	*	*	*	*	KOR	*	
CHI	Kim et al. (2022a)	2022		*	*			1	*			34	*				KOR	*	*	*	*	KOR	*	
CHI	Langer et al. (2022)	2022		*				1	*			397		*			Mult	*	*	*	*	DEU		
CHI	Langer et al. (2022)	2022		*				1	*			622		*			Mult	*	*	*	*	DEU		
CHI	Druga et al. (2022)	2022			*			35	*			34	*				USA	*	*	*	*	USA		
CHI	Jung et al. (2022)	2022		*				1	*	*		341		*				*	*	*	*	NLD		
CHI	DeVos et al. (2022)	2022			*			1	*			23	*					*	*	*	*	USA		
CHI	DeVos et al. (2022)	2022		*				14	*			22	*					*	*	*	*	USA		
CHI	DeVos et al. (2022)	2022		*			*	1	*			16	*					*	*	*	*	USA		
CHI	Mahmood et al. (2022)	2022		*				1	*	*		37	*					*	*	*	*	USA		
CHI	Lyons et al. (2022)	2022		*				1	*		*	100	*		*		USA	*	*	*	*	AUS		
CHI	Erlei et al. (2022)	2022		*				1	*		*	480	*	*				*	*	*	*	DEU	*	
CHI	Zhang and Lim (2022)	2022		*				1	*	*		14	*	*				*	*	*	*	SGP		
CHI	Zhang and Lim (2022)	2022		*				1	*	*		161	*	*				*	*	*	*	SGP		
CHI	Zhang et al. (2022b)	2022		*				1	*	*		32	*	*				*	*	*	*	SGP	*	
CHI	Zhang et al. (2022b)	2022		*				1	*	*		155	*	*				*	*	*	*	SGP	*	
CHI	Zdanowska and Taylor (2022)	2022		*				1	*	*		27	*	*				*	*	*	*	GBR		
CHI	Zdanowska and Taylor (2022)	2022		*	*			1	*	*		6	*	*				*	*	*	*	GBR		
CHI	Zdanowska and Taylor (2022)	2022		*	*			1	*	*		7	*	*				*	*	*	*	GBR		
CHI	Albayaydh and Flechais (2022)	2022		*	*			1	*	*		20	*	*			JOR	*	*	*	*	GBR		
CHI	Ma et al. (2022)	2022		*				1	*	*		67	*	*			CHN	*	*	*	*	CHN		
CHI	Ma et al. (2022)	2022		*				1	*	*		62	*	*			CHN	*	*	*	*	CHN		
CHI	Ma et al. (2022)	2022		*				1	*	*		71	*	*			CHN	*	*	*	*	CHN		
CHI	Kawakami et al. (2022)	2022		*	*			1	*	*		13	*	*			USA	*	*	*	*	USA		
CHI	Gordon et al. (2022)	2022		*				1	*	*		18	*	*				*	*	*	*	USA		
CHI	Echterhoff et al. (2022)	2022		*				1	*	*		90	*	*			Mult	*	*	*	*	USA		
CHI	Setlur and Tory (2022)	2022	*					1	*	*		30	*	*				*	*	*	*	USA		
CHI	Setlur and Tory (2022)	2022	*					1	*	*		30	*	*				*	*	*	*	USA		
CHI	Thakkar et al. (2022)	2022		*				1	*	*		46	*	*			Mult	*	*	*	*	IND	*	
CHI	Sambasivan and Veeraraghavan (2022)	2022		*	*			1	*	*		68	*	*			Mult	*	*	*	*	USA		
CHI	Zheng et al. (2022)	2022		*	*			1	*	*		7	*	*				*	*	*	*	HKG	*	
CHI	Zheng et al. (2022)	2022		*	*			1	*	*		12	*	*				*	*	*	*	HKG	*	
CHI	Carros et al. (2022)	2022		*	*			52	*	*		9	*	*			DEU	*	*	*	*	DEU	*	
CHI	Yan et al. (2022)	2022		*	*			3	*	*		15	*	*				*	*	*	*	USA	*	
CHI	Lai et al. (2022)	2022		*	*			1	*	*		234	*	*			USA	*	*	*	*	USA	*	
CHI	Zhang et al. (2022a)	2022		*	*			1	*	*		24	*	*			USA	*	*	*	*	USA		
CHI	Choi et al. (2022)	2022		*	*			1	*	*		8	*	*			KOR	*	*	*	*	KOR		
CHI	Kim et al. (2022b)	2022		*	*			1	*	*		18	*	*			USA	*	*	*	*	USA		
CHI	Cheng et al. (2022)	2022		*	*			2	*	*		13	*	*			USA	*	*	*	*	USA		
CHI	Rechkemmer and Yin (2022)	2022		*	*			1	*	*		1224	*	*			USA	*	*	*	*	USA		
CHI	Panigutti et al. (2022)	2022		*	*			1	*	*		28	*	*				*	*	*	*	ITA		
CHI	Liu et al. (2022)	2022		*	*			1	*	*		147	*	*				*	*	*	*	USA		
CHI	Liu et al. (2022)	2022		*	*			1	*	*		10	*	*			USA	*	*	*	*	USA		
FAccT	Schuff et al. (2022)	2022		*	*			1	*	*		50	*	*				*	*	*	*	DEU	*	
FAccT	Jakesch et al. (2022)	2022		*	*			1	*	*		516	*	*			USA	*	*	*	*	USA	*	
FAccT	Jakesch et al. (2022)	2022		*	*			1	*	*		607	*	*			USA	*	*	*	*	USA	*	
FAccT	Jakesch et al. (2022)	2022		*	*			1	*	*		140	*	*			USA	*	*	*	*	USA	*	
FAccT	Schoeffer et al. (2022)	2022		*	*			1	*	*		397	*	*				*	*	*	*	DEU		
FAccT	Deng et al. (2022)	2022		*	*			1	*	*		11	*	*			Mult	*	*	*	*	USA	*	
FAccT	Deng et al. (2022)	2022		*	*			1	*	*		21	*	*				*	*	*	*	USA	*	
FAccT	Ramesh et al. (2022)	2022		*	*			1	*	*		29	*	*			IND	*	*	*	*	USA	*	
FAccT	Widder et al. (2022)	2022		*	*			1	*	*		11	*	*			Mult	*	*	*	*	USA		
FAccT	Scott et al. (2022)	2022		*	*			1	*	*		29	*	*				*	*	*	*	BEL	*	
FAccT	Scott et al. (2022)	2022		*	*			1	*	*		13	*	*			Mult	*	*	*	*	BEL	*	
FAccT	Boyd (2022)	2022		*	*			1	*	*		23	*	*				*	*	*	*	USA		
FAccT	Bell et al. (2022)	2022		*	*			1	*	*		336	*	*			USA	*	*	*	*	USA	*	
FAccT	Rostamzadeh et al. (2022)	2022		*	*			1	*	*		21	*	*				*	*	*	*	CAN	*	
FAccT	Klumby et al. (2022)	2022		*	*			4	*	*		9	*	*				*	*	*	*	DEU	*	
FAccT	Costanza-Chock et al. (2022)	2022		*	*			1	*	*		152	*	*			Mult	*	*	*	*	USA		
FAccT	Costanza-Chock et al. (2022)	2022		*	*			1	*	*		10	*	*			Mult	*	*	*	*	USA		
FAccT	Stapleton et al. (2022)	2022		*	*			1	*	*		35	*	*			USA	*	*	*	*	USA	*	
FAccT	Fogliato et al. (2022)	2022	*					1	*	*		19	*	*				*	*	*	*	USA	*	

(continued on next page)

Table 3 (continued).

Conf.	Reference	Year	Lab study	Remote study	Interview study	Field study	Workshop	Duration	Qualitative	Quantitative	Mixed method	N	Domain experts	General public	Crowdworkers	HCI/ML pract.	Participant country	Gender	Race	Age	Education	Author country	Author crosscountry
FACCT	Shen et al. (2022)	2022						1	•			15	•				Mult	•				USA	•
FACCT	Smith et al. (2022)	2022			•			1	•			20				•	Mult					USA	
FACCT	Smith et al. (2022)	2022						1	•			6				•	Mult					USA	
FACCT	Smith et al. (2022)	2022		•				1	•			4				•	Mult					USA	
FACCT	Shang et al. (2022)	2022						1	•			91			•		USA	•		•	•	USA	
FACCT	Shang et al. (2022)	2022			•			1	•			12		•								USA	
FACCT	Shang et al. (2022)	2022		•				1		•		89					USA	•		•	•	USA	
FACCT	Ehsan et al. (2022)	2022			•			1	•			47		•			BGD					USA	
FACCT	Longoni et al. (2022)	2022		•				1		•		3029			•		USA	•				USA	•
FACCT	Longoni et al. (2022)	2022		•				1	•			1005					USA	•				USA	•

References

Abdul, A., Vermeulen, J., Wang, D., Lim, B.Y., Kankanalli, M., 2018. Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–18. <http://dx.doi.org/10.1145/3173574.3174156>.

Adam, A., 1998. *Artificial Knowing: Gender and the Thinking Machine*. Routledge.

Ahmed, S.I., Bidwell, N.J., Zade, H., Muralidhar, S.H., Dhreshwar, A., Karachiwala, B., Tandong, C.N., O'Neill, J., 2016. Peer-to-peer in the workplace: A view from the road. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, Association for Computing Machinery, pp. 5063–5075. <http://dx.doi.org/10.1145/2858036.2858393>.

Aitamurto, T., Chen, K., 2017. The value of crowdsourcing in public policymaking: epistemic, democratic and economic value. *Theory Pract. Legis.* 5 (1), 55–72. <http://dx.doi.org/10.1080/20508840.2017.1282665>.

Albayaydh, W.S., Flechais, I., 2022. Exploring bystanders' privacy concerns with smart homes in Jordan. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3502097>.

Alkhatib, A., 2021. To live in their utopia: Why algorithmic systems create absurd outcomes. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445740>.

Alvarado, O., Waern, A., 2018. Towards algorithmic experience: Initial efforts for social media contexts. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3173574.3173860>.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E., 2019. Guidelines for human-AI interaction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3290605.3300233>.

Andalibi, N., Buss, J., 2020. The human in emotion recognition on social media: Attitudes, outcomes, risks. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–16. <http://dx.doi.org/10.1145/3313831.3376680>.

Andrus, M., Spitzer, E., Brown, J., Xiang, A., 2021. What we can't measure, we can't understand: Challenges to demographic data procurement in the pursuit of fairness. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FACCT '21, Association for Computing Machinery, New York, NY, USA, pp. 249–260. <http://dx.doi.org/10.1145/3442188.3445888>.

Anik, A.I., Bunt, A., 2021. Data-centric explanations: Explaining training data of machine learning systems to promote transparency. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445736>.

Arnett, J.J., 2008. The neglected 95%: Why American psychology needs to become less American. *Am. Psychol.* 63 (7), 602–614. <http://dx.doi.org/10.1037/0003-066X.63.7.602>.

Ashktorab, Z., Jain, M., Liao, Q.V., Weisz, J.D., 2019. Resilient chatbots: Repair strategy preferences for conversational breakdowns. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3290605.3300484>.

Ashktorab, Z., Vitak, J., 2016. Designing cyberbullying mitigation and prevention solutions through participatory design with teenagers. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, Association for Computing Machinery, pp. 3895–3905. <http://dx.doi.org/10.1145/2858036.2858548>.

Attenberg, J.M., Ipeirotis, P.G., Provost, F., 2011. Beat the machine: Challenging workers to find the unknown unknowns. In: *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., Rahwan, I., 2018. The Moral Machine experiment. *Nature* 563 (7729), 59–64. <http://dx.doi.org/10.1038/s41586-018-0637-6>.

Bae Brandtzaeg, P.B., Skjuve, M., Kristoffer Dysthe, K.K., Følstad, A., 2021. When the social becomes non-human: Young People's perception of social support in chatbots. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445318>.

Barbosa, N.M., Chen, M., 2019. Rehumanized crowdsourcing: A labeling framework addressing bias and ethics in machine learning. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3290605.3300773>.

Bateman, S., Mandryk, R.L., Stach, T., Gutwin, C., 2011. Target assistance for subtly balancing competitive play. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '11, Association for Computing Machinery, pp. 2355–2364. <http://dx.doi.org/10.1145/1978942.1979287>.

Baughan, A., Oliveira, N., August, T., Yamashita, N., Reinecke, K., 2021. Do cross-cultural differences in visual attention patterns affect search efficiency on websites? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '21.

Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P., Vardoulakis, L.M., 2020. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3313831.3376718>.

Bell, A., Solano-Kamaiko, I., Nov, O., Stoyanovich, J., 2022. It's just not that simple: An empirical study of the accuracy-explainability trade-off in machine learning for public policy. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FACCT '22, Association for Computing Machinery, New York, NY, USA, pp. 248–266. <http://dx.doi.org/10.1145/3531146.3533090>.

Bennett, C.L., Gleason, C., Scheuerman, M.K., Bigham, J.P., Guo, A., To, A., 2021. "It's complicated": Negotiating accessibility and (mis)representation in image descriptions of race, gender, and disability. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445498>.

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., Shadbolt, N., 2018. 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3173574.3173951>.

Blake, P.R., McAuliffe, K., Corbit, J., Callaghan, T.C., Barry, O., Bowie, A., Kleutsch, L., Kramer, K.L., Ross, E., Vongsachang, H., Wrangham, R., Warneken, F., 2015. The ontology of fairness in seven societies. *Nature* 528 (7581), 258–261. <http://dx.doi.org/10.1038/nature15703>.

Bowker, G.C., Star, S.L., 2000. *Sorting Things Out: Classification and its Consequences*. MIT Press.

Boyd, K., 2022. Designing up with value-sensitive design: Building a field guide for ethical ML development. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FACCT '22, Association for Computing Machinery, New York, NY, USA, pp. 2069–2082. <http://dx.doi.org/10.1145/3531146.3534626>.

Braun, M., Mainz, A., Chadowitz, R., Pfleging, B., Alt, F., 2019. At your service: Designing voice assistant personalities to improve automotive user interfaces. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–11. <http://dx.doi.org/10.1145/3290605.3300270>.

- Brightman, M., Grotti, V., 2020. The ethics of anthropology. In: Iphofen, R. (Ed.), *Handbook of Research Ethics and Scientific Integrity*. Springer International Publishing, pp. 817–834. http://dx.doi.org/10.1007/978-3-030-16759-2_37.
- Brown, A., Chouldechova, A., Putnam-Hornstein, E., Tobin, A., Vaithianathan, R., 2019. Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3290605.3300271>.
- Brownstein, M., 2019. *Implicit Bias*. In: Zalta, E.N. (Ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2019 ed. Metaphysics Research Lab, Stanford University.
- Caine, K., 2016. Local standards for sample size at CHI. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16, Association for Computing Machinery, pp. 981–992. <http://dx.doi.org/10.1145/2858036.2858498>.
- Calders, T., Verwer, S., 2010. Three naive Bayes approaches for discrimination-free classification. *Data Min. Knowl. Discov.* 21 (2), 277–292.
- Calders, T., Žliobaitė, I., 2013. Why unbiased computational processes can lead to discriminative decision procedures. In: *Discrimination and Privacy in the Information Society*. Springer, pp. 43–57.
- Carros, F., Schwaning, I., Preussner, A., Randall, D., Wieching, R., Fitzpatrick, G., Wulf, V., 2022. Care workers making use of robots: Results of a three-month study on human-robot interaction within a care home. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517435>.
- Celis, L.E., Hays, C., Mehrotra, A., Vishnoi, N.K., 2021. The effect of the Rooney rule on implicit bias in the long term. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '21, Association for Computing Machinery, New York, NY, USA, pp. 678–689. <http://dx.doi.org/10.1145/3442188.3445930>.
- Chancellor, S., Baumer, E.P.S., De Choudhury, M., 2019. Who is the “human” in human-centered machine learning: The case of predicting mental health from social media. *Proc. ACM Hum.-Comput. Interact.* 3 (CSCW), <http://dx.doi.org/10.1145/3359249>.
- Cheng, H.-F., Stapleton, L., Kawakami, A., Sivaraman, V., Cheng, Y., Qing, D., Perer, A., Holstein, K., Wu, Z.S., Zhu, H., 2022. How child welfare workers reduce racial disparities in algorithmic decisions. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3501831>.
- Cheng, H.-F., Stapleton, L., Wang, R., Bullock, P., Chouldechova, A., Wu, Z.S.S., Zhu, H., 2021. Soliciting stakeholders' fairness notions in child maltreatment predictive systems. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445308>.
- Cheng, H.-F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F.M., Zhu, H., 2019. Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3290605.3300789>.
- Chin, H., Molefi, L.W., Yi, M.Y., 2020. Empathy is all you need: How a conversational agent should respond to verbal abuse. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376461>.
- Choi, D., Lee, U., Hong, H., 2022. “It's not wrong, but i'm quite disappointed”: Toward an inclusive algorithmic experience for content creators with disabilities. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517574>.
- Cockburn, A., Gutwin, C., Dix, A., 2018. HARK no more: On the preregistration of CHI experiments. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3173574.3173715>.
- Colley, A., Thebault-Spieker, J., Lin, A.Y., Degraen, D., Fischman, B., Häkkinen, J., Kuehl, K., Nisi, V., Nunes, N.J., Wenig, N., Wenig, D., Hecht, B., Schöning, J., 2017. The geography of Pokémon GO: Beneficial and problematic effects on places and movement. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17, Association for Computing Machinery, pp. 1179–1192. <http://dx.doi.org/10.1145/3025453.3025495>.
- Cosley, D., Lam, S.K., Albert, I., Konstan, J.A., Riedl, J., 2003. Is seeing believing? How recommender system interfaces affect users' opinions. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '03, Association for Computing Machinery, pp. 585–592. <http://dx.doi.org/10.1145/642611.642713>.
- Costanza-Chock, S., Raji, I.D., Buolamwini, J., 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, pp. 1571–1583. <http://dx.doi.org/10.1145/3531146.3533213>.
- Crenshaw, K., 1989. *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, 1989. University of Chicago Legal Forum, pp. 139–168.
- Crisan, A., Fiore-Gartland, B., 2021. Fits and starts: Enterprise use of automl and the role of humans in the loop. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445775>.
- Cryan, J., Tang, S., Zhang, X., Metzger, M., Zheng, H., Zhao, B.Y., 2020. Detecting gender stereotypes: Lexicon vs. Supervised learning methods. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20, Association for Computing Machinery, pp. 1–11. <http://dx.doi.org/10.1145/3313831.3376488>.
- Darley, J.M., Batson, C.D., 1973. “From Jerusalem to Jericho”: A study of situational and dispositional variables in helping behavior. *J. Personal. Soc. Psychol.* 27 (1), 100–108. <http://dx.doi.org/10.1037/h0034449>.
- Dell, N., Vaidyanathan, V., Medhi, I., Cutrell, E., Thies, W., 2012. “Yours is better!”: Participant response bias in HCI. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12, Association for Computing Machinery, pp. 1321–1330. <http://dx.doi.org/10.1145/2207676.2208589>.
- Deng, W.H., Nagireddy, M., Lee, M.S.A., Singh, J., Wu, Z.S., Holstein, K., Zhu, H., 2022. Exploring how machine learning practitioners (try to) use fairness toolkits. In: *2022 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 473–484. <http://dx.doi.org/10.1145/3531146.3533113>.
- Depping, A.E., Mandryk, R.L., Li, C., Gutwin, C., Vicencio-Moreira, R., 2016. How disclosing skill assistance affects play experience in a multiplayer first-person shooter game. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16, Association for Computing Machinery, pp. 3462–3472. <http://dx.doi.org/10.1145/2858036.2858156>.
- DeVito, M.A., Birnholtz, J., Hancock, J.T., 2017. Platforms, people, and perception: Using affordances to understand self-presentation on social media. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. CSCW '17, Association for Computing Machinery, pp. 740–754. <http://dx.doi.org/10.1145/2998181.2998192>.
- DeVos, A., Dhabalia, A., Shen, H., Holstein, K., Eslami, M., 2022. Toward user-driven algorithm auditing: Investigating users' strategies for uncovering harmful algorithmic behavior. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517441>.
- Dexe, J., Franke, U., 2020. Nordic lights? National AI policies for doing well by doing good. *J. Cyber Policy* 5 (3), 332–349. <http://dx.doi.org/10.1080/23738871.2020.1856160>.
- Diana, N., Stamper, J., Koedinger, K., 2020. Towards value-adaptive instruction: A data-driven method for addressing bias in argument evaluation tasks. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20, Association for Computing Machinery, pp. 1–11. <http://dx.doi.org/10.1145/3313831.3376469>.
- Difallah, D., Filatova, E., Ipeirotis, P., 2018. Demographics and dynamics of mechanical Turk workers. In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. WSDM '18, Association for Computing Machinery, pp. 135–143. <http://dx.doi.org/10.1145/3159652.3159661>.
- Dillen, N., Ilievski, M., Law, E., Nacke, L.E., Czarniecki, K., Schneider, O., 2020. Keep calm and ride along: Passenger comfort and anxiety as physiological responses to autonomous driving styles. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376247>.
- Ding, X., Jiang, Y., Qin, X., Chen, Y., Zhang, W., Qi, L., 2019. Reading face, reading health: Exploring face reading technologies for everyday health. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3290605.3300435>.
- Dolin, C., Weinschel, B., Shan, S., Hahn, C.M., Choi, E., Mazurek, M.L., Ur, B., 2018. Unpacking perceptions of data-driven inferences underlying online targeting and personalization. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3173574.3174067>.
- Dragicevic, P., 2016. Fair statistical communication in HCI. In: Robertson, J., Kaptein, M. (Eds.), *Modern Statistical Methods for HCI*. Springer International Publishing, pp. 291–330. http://dx.doi.org/10.1007/978-3-319-26633-6_13.
- Druga, S., Christoph, F.L., Ko, A.J., 2022. Family as a third space for AI literacies: How do children and parents learn about AI together? In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3502031>.
- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R., 2012. Fairness through awareness. In: *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ITCS '12, Association for Computing Machinery, pp. 214–226. <http://dx.doi.org/10.1145/2090236.2090255>.
- Echterhoff, J.M., Yarmand, M., McAuley, J., 2022. AI-moderated decision-making: Capturing and balancing anchoring bias in sequential decision tasks. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517443>.

- Echtler, F., Häußler, M., 2018. Open source, open science, and the replication crisis in HCI. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. In: CHI EA '18, Association for Computing Machinery, pp. 1–8. <http://dx.doi.org/10.1145/3170427.3188395>.
- Ehsan, U., Liao, Q.V., Muller, M., Riedl, M.O., Weisz, J.D., 2021. Expanding explainability: Towards social transparency in AI systems. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445188>.
- Ehsan, U., Singh, R., Metcalf, J., Riedl, M., 2022. The algorithmic imprint. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 1305–1317. <http://dx.doi.org/10.1145/3531146.3533186>.
- Erlei, A., Das, R., Meub, L., Anand, A., Gadiraju, U., 2022. For what it's worth: Humans overwrite their economic self-interest to avoid bargaining with AI systems. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517734>.
- Eslami, M., Krishna Kumaran, S.R., Sandvig, C., Karahalios, K., 2018. Communicating algorithmic process in online behavioral advertising. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3173574.3174006>.
- Eslami, M., Rickman, A., Vaccaro, K., Aleyas, A., Vuong, A., Karahalios, K., Hamilton, K., Sandvig, C., 2015. "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. CHI '15, Association for Computing Machinery, pp. 153–162. <http://dx.doi.org/10.1145/2702123.2702556>.
- Eslami, M., Vaccaro, K., Lee, M.K., Elazari Bar On, A., Gilbert, E., Karahalios, K., 2019. User attitudes towards algorithmic opacity and transparency in online reviewing platforms. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3290605.3300724>.
- Fan, J., Zhang, A.X., 2020. Digital juries: A civics-oriented approach to platform governance. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3313831.3376293>.
- Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., Moran, S., 2018. Falling for fake news: Investigating the consumption of news via social media. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–10. <http://dx.doi.org/10.1145/3173574.3173950>.
- Fogliato, R., Chappidi, S., Lungren, M., Fisher, P., Wilson, D., Fitzke, M., Parkinson, M., Horvitz, E., Inkpen, K., Nushi, B., 2022. Who goes first? Influences of human-AI workflow on decision making in clinical imaging. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 1362–1374. <http://dx.doi.org/10.1145/3531146.3533193>.
- Friedman, B., Nissenbaum, H., 1996. Bias in computer systems. *ACM Trans. Inf. Syst.* 14 (3), 330–347. <http://dx.doi.org/10.1145/230538.230561>.
- Gao, G., Hwang, S.Y., Culbertson, G., Fussell, S.R., Jung, M.F., 2017. Beyond information content: The effects of culture on affective grounding in instant messaging conversations. *Proc. ACM Hum.-Comput. Interact.* 1 (CSCW), <http://dx.doi.org/10.1145/3134683>.
- Geeng, C., Yee, S., Roesner, F., 2020. Fake news on facebook and Twitter: Investigating how people (don't) investigate. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3313831.3376784>.
- Gero, K.I., Ashktorab, Z., Dugan, C., Pan, Q., Johnson, J., Geyer, W., Ruiz, M., Miller, S., Millen, D.R., Campbell, M., Kumaravel, S., Zhang, W., 2020. Mental models of AI agents in a cooperative game setting. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3313831.3376316>.
- Gilad, Z., Amir, O., Levontin, L., 2021. The effects of warmth and competence perceptions on users' choice of an AI system. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3446863>.
- Gordon, M.L., Lam, M.S., Park, J.S., Patel, K., Hancock, J., Hashimoto, T., Bernstein, M.S., 2022. Jury learning: Integrating dissenting voices into machine learning models. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3502004>.
- Green, B., Chen, Y., 2019. Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. In: FAT* '19, Association for Computing Machinery, New York, NY, USA, pp. 90–99. <http://dx.doi.org/10.1145/3287560.3287563>.
- Grgic-Hlaca, N., Redmiles, E.M., Gummadi, K.P., Weller, A., 2018. Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In: Proceedings of the 2018 World Wide Web Conference. WWW '18, International World Wide Web Conferences Steering Committee, pp. 903–912. <http://dx.doi.org/10.1145/3178876.3186138>.
- Grgic-Hlaca, N., Lima, G., Weller, A., Redmiles, E.M., 2020. Dimensions of diversity in human perceptions of algorithmic fairness. <http://dx.doi.org/10.48550/ARXIV.2005.00808>, URL <https://arxiv.org/abs/2005.00808>.
- Hamidi, F., Scheuerman, M.K., Branham, S.M., 2018. Gender recognition or gender reductionism? The social implications of embedded gender recognition systems. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3173574.3173582>.
- Hanna, A., Denton, E., Smart, A., Smith-Loud, J., 2020. Towards a critical race methodology in algorithmic fairness. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. In: FAT* '20, Association for Computing Machinery, New York, NY, USA, pp. 501–512. <http://dx.doi.org/10.1145/3351095.3372826>.
- Harman, G., 1999. Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proc. Aristot. Soc.* 99, 315–331.
- Harrison, G., Hanson, J., Jacinto, C., Ramirez, J., Ur, B., 2020. An empirical study on the perceived fairness of realistic, imperfect machine learning models. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. In: FAT* '20, Association for Computing Machinery, New York, NY, USA, pp. 392–402. <http://dx.doi.org/10.1145/3351095.3372831>.
- Hastings, E.M., Alamri, A., Kuznetsov, A., Piszarczyk, C., Karahalios, K., Marinov, D., Bailey, B.P., 2020. LIFT: Integrating stakeholder voices into algorithmic team formation. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376797>.
- Henrich, J., Heine, S.J., Norenzayan, A., 2010. The weirdest people in the world? *Behav. Brain Sci.* 33 (2–3), 61–83. <http://dx.doi.org/10.1017/S0140525X0999152X>.
- Herrmann, B., Thöni, C., Gächter, S., 2008. Antisocial punishment across societies. *Science* 319 (5868), 1362–1367. <http://dx.doi.org/10.1126/science.1153808>.
- Hoffmann, A.L., 2019. Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Inf. Commun. Soc.* 22 (7), 900–915. <http://dx.doi.org/10.1080/1369118X.2019.1573912>.
- Holstein, K., Wortman Vaughan, J., Daumé, H., Dudik, M., Wallach, H., 2019. Improving fairness in machine learning systems: What do industry practitioners need? In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–16. <http://dx.doi.org/10.1145/3290605.3300830>.
- Hong, J., Lee, K., Xu, J., Kacorri, H., 2020. Crowdsourcing the perception of machine teaching. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3313831.3376428>.
- Hornbæk, K., Mottelson, A., Knibbe, J., Vogel, D., 2019. What do we mean by "interaction"? An analysis of 35 years of CHI. *ACM Trans. Comput.-Hum. Interact.* 26 (4), <http://dx.doi.org/10.1145/3325285>.
- Hornbæk, K., Sander, S.S., Bargas-Avila, J.A., Grue Simonsen, J., 2014. Is once enough? On the extent and content of replications in human-computer interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '14, Association for Computing Machinery, pp. 3523–3532. <http://dx.doi.org/10.1145/2556288.2557004>.
- Hsu, S., Li, T.W., Zhang, Z., Fowler, M., Zilles, C., Karahalios, K., 2021. Attitudes surrounding an imperfect AI autograder. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445424>.
- Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., Akkijaru, R., 2018. Touch your heart: A tone-aware chatbot for customer care on social media. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3173574.3173989>.
- van IJzendoorn, M.H., Bakermans-Kranenburg, M.J., Pannebakker, F., Out, D., 2010. In defence of situational morality: genetic, dispositional and situational determinants of children's donating to charity. *J. Moral Educ.* 39 (1), 1–20. <http://dx.doi.org/10.1080/03057240903528535>.
- Irani, L.C., Silberman, M.S., 2013. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '13, Association for Computing Machinery, pp. 611–620. <http://dx.doi.org/10.1145/2470654.2470742>.
- Irani, L., Vertesi, J., Dourish, P., Philip, K., Grinter, R.E., 2010. Postcolonial computing: A lens on design and development. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '10, Association for Computing Machinery, pp. 1311–1320. <http://dx.doi.org/10.1145/1753326.1753522>.
- Israni, E.T., 2017. When an algorithm helps send you to prison. URL <https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html>.
- Jacobs, M., He, J., F. Pradier, M., Lam, B., Ahn, A.C., McCoy, T.H., Perlis, R.H., Doshi-Velez, F., Gajos, K.Z., 2021. Designing AI for trust and collaboration in time-constrained medical decisions: A sociotechnical lens. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445385>.
- JafariNaimi, N., 2018. Our bodies in the trolley's path, or why self-driving cars must *not* be programmed to kill. *Sci. Technol. Hum. Values* 43 (2), 302–323. <http://dx.doi.org/10.1177/0162243917718942>.

- Jahanbakhsh, F., Fu, W.-T., Karahalios, K., Marinov, D., Bailey, B., 2017. You want me to work with who? Stakeholder perceptions of automated team formation in project-based courses. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17, Association for Computing Machinery, pp. 3201–3212. <http://dx.doi.org/10.1145/3025453.3026011>.
- Jakesch, M., Bućinca, Z., Amershi, S., Olteanu, A., 2022. How different groups prioritize ethical values for responsible AI. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 310–323. <http://dx.doi.org/10.1145/3531146.3533097>.
- Jakesch, M., French, M., Ma, X., Hancock, J.T., Naaman, M., 2019. AI-mediated communication: How the perception that profile text was written by AI affects trustworthiness. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3290605.3300469>.
- Jesus, S., Belém, C., Balayan, V., Bento, J., Saleiro, P., Bizarro, P., Gama, J., 2021. How can I choose an explainer? An application-grounded evaluation of post-hoc explanations. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT '21, Association for Computing Machinery, New York, NY, USA, pp. 805–815. <http://dx.doi.org/10.1145/3442188.3445941>.
- Jiang, W., Sarsenbayeva, Z., van Berkel, N., Wang, C., Yu, D., Wei, J., Goncalves, J., Kostakos, V., 2021. User trust in assisted decision-making using miniaturized near-infrared spectroscopy. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445710>.
- Jung, J.-Y., Qiu, S., Bozzon, A., Gadiraju, U., 2022. Great chain of agents: The role of metaphorical representation of agents in conversational crowdsourcing. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517653>.
- Kapania, S., Siy, O., Clapper, G., SP, A.M., Sambasivan, N., 2022. "Because AI is 100% right and safe": User attitudes and sources of AI authority in India. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517533>.
- Kasinidou, M., Kleanthous, S., Barlas, P., Otterbacher, J., 2021. I agree with the decision, but they didn't deserve this: Future developers' perception of fairness in algorithmic decisions. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT '21, Association for Computing Machinery, New York, NY, USA, pp. 690–700. <http://dx.doi.org/10.1145/3442188.3445931>.
- Kawakami, A., Sivaraman, V., Cheng, H.-F., Stapleton, L., Cheng, Y., Qing, D., Perer, A., Wu, Z.S., Zhu, H., Holstein, K., 2022. Improving human-AI partnerships in child welfare: Understanding worker practices, challenges, and desires for algorithmic decision support. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517439>.
- Keyes, O., 2018. The misgendering machines: Trans/HCI implications of automatic gender recognition. Proc. ACM Hum.-Comput. Interact. 2 (CSCW), <http://dx.doi.org/10.1145/3274357>.
- Kim, T., Kim, H., Lee, H.Y., Goh, H., Abdigapporov, S., Jeong, M., Cho, H., Han, K., Noh, Y., Lee, S.-J., Hong, H., 2022a. Prediction for retrospection: Integrating algorithmic stress prediction into personal informatics systems for college students' mental health. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517701>.
- Kim, J., Muhic, J., Robert, L.P., Park, S.Y., 2022b. Designing chatbots with black Americans with chronic conditions: Overcoming challenges against COVID-19. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3502116>.
- Kim, S., Thakur, A., Kim, J., 2020. Understanding users' perception towards automated personality detection with group-specific behavioral data. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3313831.3376250>.
- King, W.J., Ohya, J., 1996. The representation of agents: Anthropomorphism, agency, and intelligence. In: Conference Companion on Human Factors in Computing Systems. CHI '96, Association for Computing Machinery, pp. 289–290. <http://dx.doi.org/10.1145/257089.257326>.
- Kittley-Davies, J., Alqaraawi, A., Yang, R., Costanza, E., Rogers, A., Stein, S., 2019. Evaluating the effect of feedback from different computer vision processing stages: A comparative lab study. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3290605.3300273>.
- Kizilcec, R.F., 2016. How much information? Effects of transparency on trust in an algorithmic interface. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, Association for Computing Machinery, pp. 2390–2395. <http://dx.doi.org/10.1145/2858036.2858402>.
- Kjeldskov, J., Skov, M.B., 2014. Was it worth the hassle? Ten years of mobile HCI research discussions on lab and field evaluations. In: Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services. MobileHCI '14, Association for Computing Machinery, pp. 43–52. <http://dx.doi.org/10.1145/2628363.2628398>.
- Kjeldskov, J., Skov, M.B., Als, B.S., Høegh, R.T., 2004. Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In: Brewster, S., Dunlop, M. (Eds.), Mobile Human-Computer Interaction - MobileHCI 2004. Springer Berlin Heidelberg, pp. 61–73.
- Kleinberg, J., Ludwig, J., Mullainathan, S., Rambachan, A., 2018. Algorithmic fairness. AEA Pap. Proc. 108, 22–27. <http://dx.doi.org/10.1257/pandp.20181018>, URL <https://www.aeaweb.org/articles?id=10.1257/pandp.20181018>.
- Klumby, G., Draude, C., Taylor, A.S., 2022. Critical tools for machine learning: Working with intersectional critical concepts in machine learning systems design. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 1528–1541. <http://dx.doi.org/10.1145/3531146.3533207>.
- Koch, J., Lucero, A., Hegemann, L., Oulasvirta, A., 2019. May AI? Design ideation with cooperative contextual bandits. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3290605.3300863>.
- Kocielnik, R., Amershi, S., Bennett, P.N., 2019. Will you accept an imperfect AI? Exploring designs for adjusting end-user expectations of AI systems. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3290605.3300641>.
- Koeman, L., 2020. HCI/UX research: What methods do we use? URL <http://lisakoeman.nl/blog/hci-ux-research-what-methods-do-we-use/>.
- Kontogiorgos, D., van Waveren, S., Wallberg, O., Pereira, A., Leite, I., Gustafson, J., 2020. Embodiment effects in interactions with failing robots. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3313831.3376372>.
- Kou, Y., Gray, C.M., Toombs, A., Nardi, B., 2018. The politics of titling: The representation of countries in CHI papers. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. In: CHI EA '18, Association for Computing Machinery, pp. 1–10. <http://dx.doi.org/10.1145/3170427.3188409>.
- Kouchaki, M., Smith, I.H., 2014. The morning morality effect: The influence of time of day on unethical behavior. Psychol. Sci. 25 (1), 95–102. <http://dx.doi.org/10.1177/0956797613498099>, PMID: 24166855.
- Kuhlman, C., Doherty, D., Nurbekova, M., Deva, G., Phyto, Z., Schoenhagen, P.-H., VanValkenburg, M., Rundensteiner, E., Harrison, L., 2019. Evaluating preference collection methods for interactive ranking analytics. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–11. <http://dx.doi.org/10.1145/3290605.3300742>.
- Kunkel, J., Donkers, T., Michael, L., Barbu, C.-M., Ziegler, J., 2019. Let me explain: Impact of personal and impersonal explanations on trust in recommender systems. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3290605.3300717>.
- Kwet, M., 2019. Digital colonialism: US empire and the new imperialism in the Global South. Race Class 60 (4), 3–26. <http://dx.doi.org/10.1177/0306396818823172>.
- Lai, V., Carton, S., Bhatnagar, R., Liao, Q.V., Zhang, Y., Tan, C., 2022. Human-AI collaboration via conditional delegation: A case study of content moderation. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3501999>.
- Lai, V., Liu, H., Tan, C., 2020. "Why is 'Chicago' deceptive?" towards building model-driven tutorials for humans. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376873>.
- Lai, V., Tan, C., 2019. On human predictions with explanations and predictions of machine learning models: A case study on deception detection. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. In: FAT* '19, Association for Computing Machinery, New York, NY, USA, pp. 29–38. <http://dx.doi.org/10.1145/3287560.3287590>.
- Langer, M., Hunsicker, T., Feldkamp, T., König, C.J., Grgić-Hlača, N., 2022. "Look! It's a computer program! It's an algorithm! It's AI!": Does terminology affect human perceptions and evaluations of algorithmic decision-making systems? In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517527>.
- Laufer, B., Jain, S., Cooper, A.F., Kleinberg, J., Heidari, H., 2022. Four years of FAccT: A reflexive, mixed-methods analysis of research contributions, shortcomings, and future prospects. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 401–426. <http://dx.doi.org/10.1145/3531146.3533107>.
- Lazar, J., Feng, J.H., Hochheiser, H., 2017. Chapter 1 - Introduction to HCI research. In: Lazar, J., Feng, J.H., Hochheiser, H. (Eds.), Research Methods in Human Computer Interaction (Second Edition). Morgan Kaufmann, pp. 1–24. <http://dx.doi.org/10.1016/B978-0-12-805390-4.00001-7>.

- Le Bras, P., Robb, D.A., Methven, T.S., Padilla, S., Chantler, M.J., 2018. Improving user confidence in concept maps: Exploring data driven explanations. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3173574.3173978>.
- Lee, J.D., Hoffman, J.D., Hayes, E., 2004. Collision warning design to mitigate driver distraction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '04, Association for Computing Machinery, pp. 65–72. <http://dx.doi.org/10.1145/985692.985701>.
- Lee, M.K., Kim, J.T., Lizarondo, L., 2017. A human-centered approach to algorithmic services: Considerations for fair and motivating smart community service management that allocates donations to non-profit organizations. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17, Association for Computing Machinery, pp. 3365–3376. <http://dx.doi.org/10.1145/3025453.3025884>.
- Lee, M.K., Kusbit, D., Metsky, E., Dabbish, L., 2015. Working with machines: The impact of algorithmic and data-driven management on human workers. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. CHI '15, Association for Computing Machinery, pp. 1603–1612. <http://dx.doi.org/10.1145/2702123.2702548>.
- Lee, M.K., Rich, K., 2021. Who is included in human perceptions of AI?: Trust and perceived fairness around healthcare AI and cultural mistrust. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445570>.
- Leprí, B., Oliver, N., Letouzé, E., Pentland, A., Vinck, P., 2018. Fair, transparent, and accountable algorithmic decision-making processes. *Philos. Technol.* 31 (4), 611–627. <http://dx.doi.org/10.1007/s13347-017-0279-x>.
- Levy, A., Agrawal, M., Satyanarayan, A., Sontag, D., 2021. Assessing the impact of automated suggestions on decision making: Domain experts mediate model errors but take less initiative. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445522>.
- Liang, Y., Fan, H.W., Fang, Z., Miao, L., Li, W., Zhang, X., Sun, W., Wang, K., He, L., Chen, X.A., 2020. OralCam: Enabling self-examination and awareness of oral health using a smartphone camera. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376238>.
- Liao, Q.V., Gruen, D., Miller, S., 2020. Questioning the AI: Informing design practices for explainable AI user experiences. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–15. <http://dx.doi.org/10.1145/3313831.3376590>.
- Liao, Q.V., Mas-ud Hussain, M., Chandar, P., Davis, M., Khazaeni, Y., Crasso, M.P., Wang, D., Muller, M., Shami, N.S., Geyer, W., 2018. All work and no play? In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3173574.3173577>.
- Liao, M., Sundar, S.S., 2021. How should AI systems talk to users when collecting their personal information? Effects of role framing and self-referencing on human-ai interaction. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445415>.
- Liebling, D.J., Lahav, M., Evans, A., Donsbach, A., Holbrook, J., Smus, B., Boran, L., 2020. Unmet needs and opportunities for mobile translation AI. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376261>.
- Lim, B.Y., Dey, A.K., Avrahami, D., 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '09, Association for Computing Machinery, pp. 2119–2128. <http://dx.doi.org/10.1145/1518701.1519023>.
- Lima, G., Grgić-Hlača, N., Cha, M., 2021. Human perceptions on moral responsibility of AI: A case study in AI-assisted bail decision-making. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445260>.
- Linxs, S., Sturm, C., Brühlmann, F., Cassau, V., Opwis, K., Reinecke, K., 2021. How WEIRD is CHI? In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, pp. 1–18. <http://dx.doi.org/10.1145/3173574.3174156>.
- Liu, Y., Goncalves, J., Ferreira, D., Xiao, B., Hosio, S., Kostakos, V., 2014. CHI 1994–2013: Mapping two decades of intellectual progress through co-word analysis. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '14, Association for Computing Machinery, pp. 3553–3562. <http://dx.doi.org/10.1145/2556288.2556969>.
- Liu, Y., Mittal, A., Yang, D., Bruckman, A., 2022. Will AI console me when I lose my pet? Understanding perceptions of AI-mediated email writing. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517731>.
- Loepp, B., Herrmann, K., Ziegler, J., 2015. Blended recommending: Integrating interactive information filtering and algorithmic recommender techniques. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. CHI '15, Association for Computing Machinery, pp. 975–984. <http://dx.doi.org/10.1145/2702123.2702496>.
- Longoni, C., Fradkin, A., Cian, L., Pennycook, G., 2022. News from generative artificial intelligence is believed less. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 97–106. <http://dx.doi.org/10.1145/3531146.3533077>.
- Louie, R., Coenen, A., Huang, C.Z., Terry, M., Cai, C.J., 2020. Novice-AI music co-creation via AI-steering tools for deep generative models. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376739>.
- Lucic, A., Haned, H., de Rijke, M., 2020. Why does my model fail? Contrastive local explanations for retail forecasting. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. In: FAT* '20, Association for Computing Machinery, New York, NY, USA, pp. 90–98. <http://dx.doi.org/10.1145/3351095.3372824>.
- Luger, E., Sellen, A., 2016. "Like having a really bad PA": The gulf between user expectation and experience of conversational agents. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, Association for Computing Machinery, pp. 5286–5297. <http://dx.doi.org/10.1145/2858036.2858288>.
- Lyons, H., Wijenayake, S., Miller, T., Velloso, E., 2022. What's the appeal? Perceptions of review processes for algorithmic decisions. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517606>.
- Ma, N.F., Yuan, C.W., Ghafurian, M., Hanrahan, B.V., 2018. Using stakeholder theory to examine drivers' stake in uber. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3173574.3173657>.
- Ma, S., Zhou, T., Nie, F., Ma, X., 2022. Glimpse: An adaptable system for instructors to grasp student learning status in synchronous online classes. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517482>.
- MacLeod, H., Bennett, C.L., Morris, M.R., Cutrell, E., 2017. Understanding blind People's experiences with computer-generated captions of social media images. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17, Association for Computing Machinery, pp. 5988–5999. <http://dx.doi.org/10.1145/3025453.3025814>.
- Madaio, M.A., Stark, L., Wortman Vaughan, J., Wallach, H., 2020. Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3313831.3376445>.
- Mahmood, A., Fung, J.W., Won, I., Huang, C.-M., 2022. Owning mistakes sincerely: Strategies for mitigating AI errors. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517565>.
- Mallari, K., Inkpen, K., Johns, P., Tan, S., Ramesh, D., Kamar, E., 2020. Do I look like a criminal? Examining how race presentation impacts human judgement of recidivism. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376257>.
- Marathe, M., Toyama, K., 2018. Semi-automated coding for qualitative research: A user-centered inquiry and initial prototypes. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3173574.3173922>.
- Marcinkowski, F., Kieslich, K., Starke, C., Lünich, M., 2020. Implications of AI (un-)fairness in higher education admissions: The effects of perceived AI (un-)fairness on exit, voice and organizational reputation. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. In: FAT* '20, Association for Computing Machinery, New York, NY, USA, pp. 122–130. <http://dx.doi.org/10.1145/3351095.3372867>.
- Maulsby, D., Greenberg, S., Mander, R., 1993. Prototyping an intelligent agent through Wizard of Oz. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. CHI '93, Association for Computing Machinery, pp. 277–284. <http://dx.doi.org/10.1145/169059.169215>.
- McCormack, J., Gifford, T., Hutchings, P., Llano Rodriguez, M.T., Yee-King, M., d'Inverno, M., 2019. In a silent way: Communication between AI and improvising musicians beyond sound. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–11. <http://dx.doi.org/10.1145/3290605.3300268>.
- Mendez, G., Galárraga, L., Chiluita, K., 2021. Showing academic performance predictions during term planning: Effects on students' decisions, behaviors, and preferences. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445718>.

- Miceli, M., Yang, T., Naudts, L., Schuessler, M., Serbanescu, D., Hanna, A., 2021. Documenting computer vision datasets: An invitation to reflexive data practices. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT '21, Association for Computing Machinery, New York, NY, USA, pp. 161–172. <http://dx.doi.org/10.1145/3442188.3445880>.
- Moritz, D., Fisher, D., Ding, B., Wang, C., 2017. Trust, but verify: Optimistic visualizations of approximate queries for exploring big data. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17, Association for Computing Machinery, pp. 2904–2915. <http://dx.doi.org/10.1145/3025453.3025456>.
- Moshontz, H., Campbell, L., Ebersole, C.R., IJzerman, H., Urry, H.L., Forscher, P.S., Grahe, J.E., McCarthy, R.J., Musser, E.D., Antfolk, J., Castille, C.M., Evans, T.R., Fiedler, S., Flake, J.K., Forero, D.A., Janssen, S.M.J., Keene, J.R., Protzko, J., Aczel, B., Álvarez Solas, S., Ansari, D., Awlia, D., Baskin, E., Batres, C., Borrascavevara, M.L., Brick, C., Chandel, P., Chatard, A., Chopik, W.J., Clarence, D., Coles, N.A., Corker, K.S., Dixon, B.J.W., Dranseika, V., Dunham, Y., Fox, N.W., Gardiner, G., Garrison, S.M., Gill, T., Hahn, A.C., Jaeger, B., Kačmár, P., Kaminiski, G., Kanske, P., Kekacs, Z., Kline, M., Koehn, M.A., Kujur, P., Levitan, C.A., Miller, J.K., Okan, C., Olsen, J., Oviedo-Trespalacios, O., Özdoğru, A.A., Pande, B., Parganiha, A., Parveen, N., Pfuhl, G., Pradhan, S., Ropovik, I., Rule, N.O., Saunders, B., Schei, V., Schmidt, K., Singh, M.M., Sirota, M., Steltenpohl, C.N., Stieger, S., Storage, D., Sullivan, G.B., Szabelska, A., Tamnes, C.K., Vadillo, M.A., Valentova, J.V., Vanpaemel, W., Varella, M.A.C., Vergauwe, E., Verschoor, M., Vianello, M., Voracek, M., Williams, G.P., Wilson, J.P., Zickfeld, J.H., Arnal, J.D., Aydin, B., Chen, S.-C., DeBruine, L.M., Fernandez, A.M., Horstmann, K.T., Isager, P.M., Jones, B., Kapucu, A., Lin, H., Mensink, M.C., Navarrete, G., Silan, M.A., Chartier, C.R., 2018. The psychological science accelerator: Advancing psychology through a distributed collaborative network. *Adv. Methods Pract. Psychol. Sci.* 1 (4), 501–515. <http://dx.doi.org/10.1177/2515245918797607>.
- Mustafaraj, E., Lurie, E., Devine, C., 2020. The case for voter-centered audits of search engines during political elections. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. In: FAT* '20, Association for Computing Machinery, New York, NY, USA, pp. 559–569. <http://dx.doi.org/10.1145/3351095.3372835>.
- Musthag, M., Raji, A., Ganesan, D., Kumar, S., Shiffman, S., 2011. Exploring micro-incentive strategies for participant compensation in high-burden studies. In: Proceedings of the 13th International Conference on Ubiquitous Computing. UbiComp '11, Association for Computing Machinery, pp. 435–444. <http://dx.doi.org/10.1145/2030112.2030170>.
- Nakao, Y., Stumpf, S., Ahmed, S., Naseer, A., Strappelli, L., 2022. Toward involving end-users in interactive human-in-the-loop AI fairness. *ACM Trans. Interact. Syst.* 12 (3), <http://dx.doi.org/10.1145/3514258>.
- Nass, C., Moon, Y., Carney, P., 1999. Are people polite to computers? Responses to computer-based interviewing Systems I. *J. Appl. Soc. Psychol.* 29 (5), 1093–1109. <http://dx.doi.org/10.1111/j.1559-1816.1999.tb00142.x>.
- Nielsen, C.M., Overgaard, M., Pedersen, M.B., Stage, J., Stenild, S., 2006. It's worth the hassle! the added value of evaluating the usability of mobile systems in the field. In: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles. NordiCHI '06, Association for Computing Machinery, pp. 272–280. <http://dx.doi.org/10.1145/1182475.1182504>.
- Noriega-Campero, A., Garcia-Bulle, B., Cantu, L.F., Bakker, M.A., Tejerina, L., Pentland, A., 2020. Algorithmic targeting of social policies: Fairness, accuracy, and distributed governance. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. In: FAT* '20, Association for Computing Machinery, New York, NY, USA, pp. 241–251. <http://dx.doi.org/10.1145/3351095.3375784>.
- Norwegian Ministry of Local Government and Modernisation, 2019. National strategy for artificial intelligence. <https://www.regjeringen.no/en/dokumenter/nasjonalstrategi-for-kunstig-intelligens/id2685594>.
- Ogbonnaya-Ogburu, I.F., Smith, A.D., To, A., Toyama, K., 2020. Critical race theory for HCI. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–16. <http://dx.doi.org/10.1145/3313831.3376392>.
- Oh, C., Choi, J., Lee, S., Park, S., Kim, D., Song, J., Kim, D., Lee, J., Suh, B., 2020. Understanding user perception of automated news generation system. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376811>.
- Okolo, C.T., Kamath, S., Dell, N., Vashistha, A., 2021. "It cannot do all of my work": Community health worker perceptions of AI-enabled mobile health applications in rural India. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445420>.
- Panigutti, C., Beretta, A., Giannotti, F., Pedreschi, D., 2022. Understanding the impact of explanations on advice-taking: A user study for AI-based clinical decision support systems. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3502104>.
- Paolacci, G., Chandler, J., Ipeirotis, P.G., 2010. Running experiments on Amazon Mechanical Turk. *Judgm. Decis. Mak.* 5 (5), 411–419.
- Park, H., Ahn, D., Hosanagar, K., Lee, J., 2021. Human-AI interaction in human resource management: Understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445304>.
- Park, H., Ahn, D., Hosanagar, K., Lee, J., 2022. Designing fair AI in human resource management: Understanding tensions surrounding algorithmic evaluation and envisioning stakeholder-centered solutions. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517672>.
- Pashler, H., Wagenmakers, E.-J., 2012. Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspect. Psychol. Sci.* 7 (6), 528–530. <http://dx.doi.org/10.1177/1745691612465253>.
- Passi, S., Barocas, S., 2019. Problem formulation and fairness. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. In: FAT* '19, Association for Computing Machinery, New York, NY, USA, pp. 39–48. <http://dx.doi.org/10.1145/3287560.3287567>.
- Pater, J., Coupe, A., Pfafman, R., Phelan, C., Toscos, T., Jacobs, M., 2021. Standardizing reporting of participant compensation in HCI: A systematic literature review and recommendations for the field. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445734>.
- Pierson, E., 2017. Demographics and discussion influence views on algorithmic fairness. <http://dx.doi.org/10.48550/ARXIV.1712.09124>, URL <https://arxiv.org/abs/1712.09124>.
- Pohl, H., Mottelson, A., 2019. How we guide, write, and cite at CHI. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. In: CHI EA '19, Association for Computing Machinery, pp. 1–11. <http://dx.doi.org/10.1145/3290607.3310429>.
- Rader, E., Cotter, K., Cho, J., 2018. Explanations as mechanisms for supporting algorithmic transparency. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3173574.3173677>.
- Rader, E., Gray, R., 2015. Understanding user beliefs about algorithmic curation in the facebook news feed. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. CHI '15, Association for Computing Machinery, pp. 173–182. <http://dx.doi.org/10.1145/2702123.2702174>.
- Rahim, Z., 2020. Algorithms promised efficiency. But they've worsened inequality. URL <https://edition.cnn.com/2020/08/23/tech/algorithms-bias-inequality-intl-gbr/index.html>.
- Rahman, R., Rahman, M.R., Tripto, N.I., Ali, M.E., Apon, S.H., Shahriyar, R., 2021. AdolescentBot: Understanding opportunities for chatbots in combating adolescent sexual and reproductive health problems in Bangladesh. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445694>.
- Ramesh, D., Kameswaran, V., Wang, D., Sambasivan, N., 2022. How platform-user power relations shape algorithmic accountability: A case study of instant loan platforms and financially stressed users in India. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 1917–1928. <http://dx.doi.org/10.1145/3531146.3533237>.
- Rechkemmer, A., Yin, M., 2022. When confidence meets accuracy: Exploring the effects of multiple performance indicators on trust in machine learning models. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3501967>.
- Rietz, T., Maedche, A., 2021. Cody: An AI-based system to semi-automate coding for qualitative research. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445591>.
- Robertson, S., Nguyen, T., Salehi, N., 2021a. Modeling assumptions clash with the real world: Transparency, equity, and community challenges for student assignment algorithms. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445748>.
- Robertson, R.E., Olteanu, A., Diaz, F., Shokouhi, M., Bailey, P., 2021b. "I can't reply with that": Characterizing problematic email reply suggestions. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445557>.
- Rogers, Y., Connelly, K., Tedesco, L., Hazlewood, W., Kurtz, A., Hall, R.E., Hursey, J., Toscos, T., 2007. Why it's worth the hassle: The value of in-situ studies when designing ubicomp. In: Krumm, J., Abowd, G.D., Seneviratne, A., Strang, T. (Eds.), *UbiComp 2007: Ubiquitous Computing*. Springer Berlin Heidelberg, pp. 336–353.
- Rostamzadeh, N., Mincu, D., Roy, S., Smart, A., Wilcox, L., Pushkarna, M., Schrouff, J., Amironesei, R., Moorosi, N., Heller, K., 2022. Healthsheet: Development of a transparency artifact for health datasets. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 1943–1961. <http://dx.doi.org/10.1145/3531146.3533239>.

- Roy, Q., Zhang, F., Vogel, D., 2019. Automation accuracy is good, but high controllability may be better. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–8. <http://dx.doi.org/10.1145/3290605.3300750>.
- Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., Prabhakaran, V., 2021a. Re-imagining algorithmic fairness in India and beyond. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT '21, Association for Computing Machinery, New York, NY, USA, pp. 315–328. <http://dx.doi.org/10.1145/3442188.3445896>.
- Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., Aroyo, L.M., 2021b. “Everyone wants to do the model work, not the data work”: Data cascades in high-stakes AI. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445518>.
- Sambasivan, N., Veeraraghavan, R., 2022. The deskilling of domain expertise in AI development. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517578>.
- Samrose, S., McDuff, D., Sim, R., Suh, J., Rowan, K., Hernandez, J., Rintel, S., Moynihan, K., Czerwinski, M., 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, <http://dx.doi.org/10.1145/3411764.3445615>.
- Saxena, N.A., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D.C., Liu, Y., 2019. How do fairness definitions fare? Examining public attitudes towards algorithmic definitions of fairness. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. AIES '19, Association for Computing Machinery, pp. 99–106. <http://dx.doi.org/10.1145/3306618.3314248>.
- Schaekermann, M., Beaton, G., Sanoubari, E., Lim, A., Larson, K., Law, E., 2020. Ambiguity-aware AI assistants for medical data analysis. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3313831.3376506>.
- Scheuerman, M.K., Paul, J.M., Brubaker, J.R., 2019. How computers see gender: An evaluation of gender classification in commercial facial analysis services. Proc. ACM Hum.-Comput. Interact. 3 (CSCW), <http://dx.doi.org/10.1145/3359246>.
- Schlesinger, A., O'Hara, K.P., Taylor, A.S., 2018. Let's talk about race: Identity, chatbots, and AI. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3173574.3173889>.
- Schneider, T., Hois, J., Rosenstein, A., Ghellal, S., Theofanou-Fülbier, D., Gerlicher, A.R., 2021. Explain yourself! transparency for positive UX in autonomous driving. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3446647>.
- Schneider, H., Wayrauther, J., Hassib, M., Butz, A., 2019. Communicating uncertainty in fertility prognosis. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–11. <http://dx.doi.org/10.1145/3290605.3300391>.
- Schoeffer, J., Kuehl, N., Machowski, Y., 2022. “There is not enough information”: On the effects of explanations on perceptions of informational fairness and trustworthiness in automated decision-making. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 1616–1628. <http://dx.doi.org/10.1145/3531146.3533218>.
- Schuff, H., Jacovi, A., Adel, H., Goldberg, Y., Vu, N.T., 2022. Human interpretation of saliency-based explanation over text. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 611–636. <http://dx.doi.org/10.1145/3531146.3533127>.
- Schuß, M., Wintersberger, P., Riemer, A., 2021. Let's share a ride into the future: A qualitative study comparing hypothetical implementation scenarios of automated vehicles. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445609>.
- Scott, K.M., Wang, S.M., Miceli, M., Delobelle, P., Sztandar-Sztanderska, K., Berendt, B., 2022. Algorithmic tools in public employment services: Towards a jobseeker-centric perspective. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 2138–2148. <http://dx.doi.org/10.1145/3531146.3534631>.
- Setlur, V., Torry, M., 2022. How do you converse with an analytical chatbot? Revisiting gricean maxims for designing analytical conversational behavior. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3501972>.
- Shang, R., Feng, K.J.K., Shah, C., 2022. Why am I not seeing it? Understanding users' needs for counterfactual explanations in everyday recommendations. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 1330–1340. <http://dx.doi.org/10.1145/3531146.3533189>.
- Shen, H., Deng, W.H., Chattopadhyay, A., Wu, Z.S., Wang, X., Zhu, H., 2021. Value cards: An educational toolkit for teaching social impacts of machine learning through deliberation. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT '21, Association for Computing Machinery, New York, NY, USA, pp. 850–861. <http://dx.doi.org/10.1145/3442188.3445971>.
- Shen, H., Wang, L., Deng, W.H., Brusse, C., Velgersdijk, R., Zhu, H., 2022. The model card authoring toolkit: Toward community-centered, deliberation-driven AI design. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 440–451. <http://dx.doi.org/10.1145/3531146.3533110>.
- Skirpan, M.W., Yeh, T., Fiesler, C., 2018. What's at stake: Characterizing risk perceptions of emerging technologies. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3173574.3173644>.
- Smith, J.J., Amershi, S., Barocas, S., Wallach, H., Wortman Vaughan, J., 2022. REAL ML: Recognizing, exploring, and articulating limitations of machine learning research. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 587–597. <http://dx.doi.org/10.1145/3531146.3533122>.
- Smith, C.E., Yu, B., Srivastava, A., Halfaker, A., Terveen, L., Zhu, H., 2020. Keeping community in the loop: Understanding wikipedia stakeholder values for machine learning-based systems. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3313831.3376783>.
- Smith-Renner, A., Fan, R., Birchfield, M., Wu, T., Boyd-Graber, J., Weld, D.S., Findlater, L., 2020. No explainability without accountability: An empirical study of explanations and feedback in interactive ML. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376624>.
- Solomon, J., 2014. Customization bias in decision support systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '14, Association for Computing Machinery, pp. 3065–3074. <http://dx.doi.org/10.1145/2556288.2557211>.
- Spiel, K., Gerling, K., 2021. The purpose of play: How HCI games research fails neurodivergent populations. ACM Trans. Comput.-Hum. Interact. 28 (2), <http://dx.doi.org/10.1145/3432245>.
- Stapleton, L., Lee, M.H., Qing, D., Wright, M., Chouldechova, A., Holstein, K., Wu, Z.S., Zhu, H., 2022. Imagining new futures beyond predictive systems in child welfare: A qualitative study with impacted stakeholders. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 1162–1177. <http://dx.doi.org/10.1145/3531146.3533177>.
- Stone, A.A., Kessler, R.C., Haythomthwatte, J.A., 1991. Measuring daily events and experiences: Decisions for the researcher. J. Personal. 59 (3), 575–607. <http://dx.doi.org/10.1111/j.1467-6494.1991.tb00260.x>.
- Sun, D., Feng, Z., Chen, Y., Wang, Y., Zeng, J., Yuan, M., Pong, T.-C., Qu, H., 2020. DFSeer: A visual analytics approach to facilitate model selection for demand forecasting. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376866>.
- Sun, K., Mhaidli, A.H., Watel, S., Brooks, C.A., Schaub, F., 2019. It's my data! tensions among stakeholders of a learning analytics dashboard. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3290605.3300824>.
- Sundar, S.S., Kim, J., 2019. Machine heuristic: When we trust computers more than humans with our personal information. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–9. <http://dx.doi.org/10.1145/3290605.3300768>.
- Swearngin, A., Dontcheva, M., Li, W., Brandt, J., Dixon, M., Ko, A.J., 2018. Rewrite: Interface design assistance from examples. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3173574.3174078>.
- Tahir, R., Batool, B., Jamshed, H., Jameel, M., Anwar, M., Ahmed, F., Zaffar, M.A., Zaffar, M.F., 2021. Seeing is believing: Exploring perceptual differences in DeepFake videos. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445699>.
- Thakkar, D., Ismail, A., Kumar, P., Hanna, A., Sambasivan, N., Kumar, N., 2022. When is machine learning data good?: Valuing in public health datafication. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3501868>.
- Thakkar, D., Kumar, N., Sambasivan, N., 2020. Towards an AI-powered future that works for vocational workers. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376674>.
- The White House - Office of Science and Technology Policy, 2020. American AI initiative - One year annual report. <https://www.nitrd.gov/nitrdgroups/images/c1/American-AI-Initiative-One-Year-Annual-Report.pdf>.

- Tolmeijer, S., Christen, M., Kandul, S., Kneer, M., Bernstein, A., 2022. Capable but amoral? Comparing AI and human expert collaboration in ethical decision making. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517732>.
- Tsai, C.-H., You, Y., Gui, X., Kou, Y., Carroll, J.M., 2021. Exploring and promoting diagnostic transparency and explainability in online symptom checkers. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445101>.
- Tullio, J., Dey, A.K., Chalecki, J., Fogarty, J., 2007. How it works: A field study of non-technical users interacting with an intelligent system. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '07, Association for Computing Machinery, pp. 31–40. <http://dx.doi.org/10.1145/1240624.1240630>.
- Uhde, A., Schlicker, N., Wallach, D.P., Hassenzehl, M., 2020. Fairness and decision-making in collaborative shift scheduling systems. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376656>.
- Ur, B., Alfieri, F., Aung, M., Bauer, L., Christin, N., Colnago, J., Cranor, L.F., Dixon, H., Emami Naeini, P., Habib, H., Johnson, N., Melicher, W., 2017. Design and evaluation of a data-driven password meter. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17, Association for Computing Machinery, pp. 3775–3786. <http://dx.doi.org/10.1145/3025453.3026050>.
- Vaccaro, K., Huang, D., Eslami, M., Sandvig, C., Hamilton, K., Karahalios, K., 2018. The illusion of control: Placebo effects of control settings. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3173574.3173590>.
- van Berkel, N., Goncalves, J., Hettiachchi, D., Wijenayake, S., Kelly, R.M., Kostakos, V., 2019. Crowdsourcing perceptions of fair predictors for machine learning: A recidivism case study. *Proc. ACM Hum.-Comput. Interact.* 3 (CSCW), <http://dx.doi.org/10.1145/3359130>.
- van Berkel, N., Goncalves, J., Russo, D., Hosio, S., Skov, M.B., 2021. Effect of information presentation on fairness perceptions of machine learning predictors. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445365>.
- van Berkel, N., Papachristos, E., Giachanou, A., Hosio, S., Skov, M.B., 2020. A systematic assessment of national artificial intelligence policies: Perspectives from the Nordics and beyond. In: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society. NordiCHI '20, Association for Computing Machinery, <http://dx.doi.org/10.1145/3419249.3420106>.
- van Berkel, N., Tag, B., Goncalves, J., Hosio, S., 2022. Human-centred artificial intelligence: a contextual morality perspective. *Behav. Inf. Technol.* 41 (3), 502–518. <http://dx.doi.org/10.1080/0144929X.2020.1818828>.
- Vashistha, A., Okeke, F., Anderson, R., Dell, N., 2018. "You can always do better!": The impact of social proof on participant response bias. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3173574.3174126>.
- Vaughan, J.W., 2017. Making better use of the crowd: How crowdsourcing can advance machine learning research. *J. Mach. Learn. Res.* 18 (1), 7026–7071.
- Veale, M., Van Kleef, M., Binns, R., 2018. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3173574.3174014>.
- Verame, J.K.M., Costanza, E., Ramchurn, S.D., 2016. The effect of displaying system confidence information on the usage of autonomous systems for non-specialist applications: A lab study. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, Association for Computing Machinery, pp. 4908–4920. <http://dx.doi.org/10.1145/2858036.2858369>.
- Verma, N., Dombrowski, L., 2018. Confronting social criticisms: Challenges when adopting data-driven policing strategies. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3173574.3174043>.
- Verma, S., Rubin, J., 2018. Fairness definitions explained. In: 2018 IEEE/ACM International Workshop on Software Fairness (FairWare). pp. 1–7. <http://dx.doi.org/10.23919/FAIRWARE.2018.8452913>.
- Völkel, S.T., Haeusschmid, R., Werner, A., Hussmann, H., Butz, A., 2020. How to trick AI: Users' strategies for protecting themselves from automatic personality assessment. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–15. <http://dx.doi.org/10.1145/3313831.3376877>.
- Wacharamanotham, C., Eisenring, L., Haroz, S., Ehtler, F., 2020. Transparency of CHI research artifacts: Results of a self-reported survey. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3313831.3376448>.
- Wang, D., Andres, J., Weisz, J.D., Oduor, E., Dugan, C., 2021a. AutoDS: Towards human-centered automation of data science. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445526>.
- Wang, R., Harper, F.M., Zhu, H., 2020. Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3313831.3376813>.
- Wang, Q., Ming, Y., Jin, Z., Shen, Q., Liu, D., Smith, M.J., Veeramachaneni, K., Qu, H., 2019a. ATMSeer: Increasing transparency and controllability in automated machine learning. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3290605.3300911>.
- Wang, D., Wang, L., Zhang, Z., Wang, D., Zhu, H., Gao, Y., Fan, X., Tian, F., 2021b. "Brilliant AI doctor" in rural clinics: Challenges in AI-powered clinical decision support system deployment. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445432>.
- Wang, D., Yang, Q., Abdul, A., Lim, B.Y., 2019b. Designing theory-driven user-centric explainable AI. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–15. <http://dx.doi.org/10.1145/3290605.3300831>.
- Warshaw, J., Matthews, T., Whittaker, S., Kau, C., Bengualid, M., Smith, B.A., 2015. Can an algorithm know the "real you"? Understanding People's reactions to hyper-personal analytics systems. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. CHI '15, Association for Computing Machinery, pp. 797–806. <http://dx.doi.org/10.1145/2702123.2702274>.
- Wexler, J., Pushkarna, M., Bolukbasi, T., Wattenberg, M., Viégas, F., Wilson, J., 2020. The what-if tool: Interactive probing of machine learning models. *IEEE Trans. Vis. Comput. Graphics* 26 (1), 56–65.
- Widder, D.G., Dabbish, L., Herbsleb, J.D., Holloway, A., Davidoff, S., 2021. Trust in collaborative automation in high stakes software engineering work: A case study at NASA. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445650>.
- Widder, D.G., Nafus, D., Dabbish, L., Herbsleb, J., 2022. Limits and possibilities for "Ethical AI" in open source: A study of deepfakes. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. FAccT '22, Association for Computing Machinery, New York, NY, USA, pp. 2035–2046. <http://dx.doi.org/10.1145/3531146.3533779>.
- Williamson, V., 2016. On the ethics of crowdsourced research. *PS: Political Sci. Politics* 49 (1), 77–81. <http://dx.doi.org/10.1017/S104909651500116X>.
- Wiseman, S., Cox, A.L., Gould, S.J., Brumby, D.P., 2017. Exploring the effects of non-monetary reimbursement for participants in HCI research. *Hum. Comput.*
- Woodruff, A., Fox, S.E., Rouso-Schindler, S., Warshaw, J., 2018. A qualitative exploration of perceptions of algorithmic fairness. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–14. <http://dx.doi.org/10.1145/3173574.3174230>.
- Wu, A.Y., Munteanu, C., 2018. Understanding older users' acceptance of wearable interfaces for sensor-based fall risk assessment. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3173574.3173693>.
- Xie, Y., Chen, M., Kao, D., Gao, G., Chen, X.A., 2020. CheXplain: Enabling physicians to explore and understand data-driven, AI-enabled medical imaging analysis. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376807>.
- Xu, K., Guo, S., Cao, N., Gotz, D., Xu, A., Qu, H., Yao, Z., Chen, Y., 2018. ECGLens: Interactive visual exploration of large scale ECG data for arrhythmia detection. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3173574.3174237>.
- Yamamoto, Y., Tanaka, K., 2011. Enhancing credibility judgment of web search results. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '11, Association for Computing Machinery, pp. 1235–1244. <http://dx.doi.org/10.1145/1978942.1979126>.
- Yan, J.N., Gu, Z., Lin, H., Rzeszotarski, J.M., 2020. Silva: Interactively assessing machine learning fairness using causality. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, pp. 1–13. <http://dx.doi.org/10.1145/3313831.3376447>.
- Yan, Z., Wu, Y., Zhang, Y., Chen, X.A., 2022. EmoGlass: An end-to-end AI-enabled wearable platform for enhancing self-awareness of emotional health. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3501925>.
- Yang, Q., Zimmerman, J., Steinfeld, A., Carey, L., Antaki, J.F., 2016. Investigating the heart pump implant decision process: Opportunities for decision support tools to help. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, Association for Computing Machinery, pp. 4477–4488. <http://dx.doi.org/10.1145/2858036.2858373>.
- Yin, M., Wortman Vaughan, J., Wallach, H., 2019. Understanding the effect of accuracy on trust in machine learning models. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, pp. 1–12. <http://dx.doi.org/10.1145/3290605.3300509>.

- You, Y., Kou, Y., Ding, X., Gui, X., 2021. The medical authority of AI: A study of AI-enabled consumer-facing health technology. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445657>.
- Yudkin, D.A., Gantman, A.P., Hofmann, W., Quoidbach, J., 2019. Moral values change importance in the presence of close others. <http://dx.doi.org/10.31234/osf.io/tcq65>.
- Zdanowska, S., Taylor, A.S., 2022. A study of UX practitioners roles in designing real-world, enterprise ML systems. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517607>.
- Zehring, R., Singhal, A., Correll, M., Battle, L., 2021. Vis ex machina: An analysis of trust in human versus algorithmically generated visualization recommendations. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, <http://dx.doi.org/10.1145/3411764.3445195>.
- Zhang, A., Boltz, A., Wang, C.W., Lee, M.K., 2022a. Algorithmic management reimaged for workers and by workers: Centering worker well-being in gig work. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3501866>.
- Zhang, W., Dimiccoli, M., Lim, B.Y., 2022b. Debiased-CAM to mitigate image perturbations with faithful visual explanations of machine learning. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517522>.
- Zhang, Y., Liao, Q.V., Bellamy, R.K.E., 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. In: FAT* '20, Association for Computing Machinery, New York, NY, USA, pp. 295–305. <http://dx.doi.org/10.1145/3351095.3372852>.
- Zhang, W., Lim, B.Y., 2022. Towards relatable explainable AI with the perceptual process. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3501826>.
- Zheng, C., Wang, D., Wang, A.Y., Ma, X., 2022. Telling stories from computational notebooks: AI-assisted presentation slides creation for presenting data science work. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, <http://dx.doi.org/10.1145/3491102.3517615>.