Biometric Mirror: Exploring Values and Attitudes towards Facial Analysis and Automated Decision-Making

Niels Wouters[1], **Ryan Kelly**[1], **Eduardo Velloso**[1], **Katrin Wolf**[2, 1],
Hasan Shahid Ferdous[1], **Joshua Newn**[1], **Zaher Joukhadar**[1], **Frank Vetere**[1]
[1] School of Computing and Information Systems, University of Melbourne, Australia
{niels.wouters, ryan.kelly, eduardo.velloso, hasan.ferdous,
joshua.newn, zaher.joukhadar, f.vetere}@unimelb.edu.au
[2] Hamburg University of Applied Sciences, Hamburg, Germany — katrin.wolf@haw-hamburg.de

ABSTRACT
Facial analysis applications are increasingly being applied to inform decision-making processes. However, as global reports of unfairness emerge, governments, academia and industry have recognized the ethical limitations and societal implications of this technology. Alongside initiatives that aim to formulate ethical frameworks, we believe that the public should be invited to participate in the debate. In this paper, we discuss *Biometric Mirror*, a case study that explored opinions about the ethics of an emerging technology. The interactive application distinguished demographic and psychometric information from people's facial photos and presented speculative scenarios with potential consequences based on their results. We analyzed the interactions with Biometric Mirror and media reports covering the study. Our findings demonstrate the nature of public opinion about the technology's possibilities, reliability, and privacy implications. Our study indicates an opportunity for case study-based digital ethics research, and we provide practical guidelines for designing future studies.

Author Keywords
Digital ethics, machine learning, facial analysis, field study, fairness, accountability, transparency, public space.

CCS Concepts
•**Human-centered computing** → **HCI theory, concepts and models;** Field studies;

INTRODUCTION
New and emerging technologies have the potential to impact all aspects of human life, yet their potential benefits may be outweighed by the societal concerns they raise [1, 2, 20, 71, 88]. For instance, while big data analytics can help drive business decisions [61], it also enables social media data to be de-anonymized, compromising users' privacy [11, 17]. Similarly, advances in computer graphics produce more realistic movies [54], and yet are also used to produce 'deepfakes' that deliberately misinform the public [37]. Though these technologies may have been created to yield a positive impact in the world, their use may also lead to unforeseen consequences that impact everyday life in ways that are negative and unwanted.

Particular concerns have been raised about the use of facial analysis technology. The technology provides opportunities to infer personal characteristics from faces in photos, video recording, and camera feeds. Inferences include gender, age, emotion and race [63, 65], but this set is continuously expanding. As the accuracy of these classifiers improves, these systems are being used to automate decision-making processes that affect access to health care, real estate, financial services, the judicial system and many more [44]. Given the lack of transparency in the application of this technology, there rarely is an opportunity for the public to question and critique the appropriateness of automated decisions [71, 88]. The significance of the technology's detrimental ethical implications is made apparent in a wealth of recent academic studies [2, 55, 58, 69, 70, 71] and news reports [4, 16, 77, 85].

The current hype surrounding facial analysis technology comes with a lack of public understanding of its implications, thereby complicating a balanced discussion between the development community, the public, and those who adopt the technology. We specifically want to understand opinions about the technology better and explore how to collect feedback about (un)ethical use cases, such as for decision-making purposes. Hence, in an effort to better understand opinions towards the technology and its impact, we developed *Biometric Mirror*. The interactive application is a provocative demonstrator that enabled people to have their face photographed and to view the inferences about their demographic and psychometric characteristics made by an inherently flawed and biased machine learning model. Interactions concluded with a speculative scenario of a decision based on the inferences and a prompt for personal reflection on the implications of an unethical decision-making process.

In this paper, we document the opinions that emerged as people witnessed their facial analysis and automated decision. Our findings reveal several opportunities to increase public awareness about the potential implications of facial analysis. This analysis and the documentation of the design process il-

DIS'19, June 23–28, 2019, San Diego, CA, USA
 ISBN 978-1-4503-5850-7/19/06…$15.00
DOI: https://doi.org/10.1145/3322276.3322304

lustrate how uncharted thought experiments can be conducted in a naturalistic setting while balancing users' well-being. Our study indicates an opportunity for case study-based digital ethics research, and we provide practical guidelines.

RELATED WORK

There is a long-standing tradition within the HCI community of conducting research that is designed to understand human behavior and to inform the design and functionality of new technologies. The expertise that has been built and shared over the years highlights a particular opportunity for the HCI community to take a leading role in studying the ethical implications of emerging technologies. We set the scene by reviewing three relevant approaches in HCI research related to (1) studying human behaviors in natural environments, (2) gathering insights into ethics and values, and (3) seeking confrontation to enable public discourse.

Ethics of Facial Analysis Technology

Industry and authorities are showing increasing interest in facial analysis technology, developing new ways to deliver insights based on the characteristics that can be distinguished from human faces. Common applications include out-of-home advertising that selects information to suit the profile of the audience [60, 63, 67]. There is broad consensus that research in this domain is progressing steadily, and that its impact on society is likely to increase. However, some applications can now also automate decisions based on perceived human behaviors, personality, sexuality and race from facial characteristics, predicting the unethical use of the technology [32, 36, 49, 86].

Lack of transparency about the inner workings of the technology and absence of opportunities for the public to question the validity of inferences necessitated a discussion on the ethics of the technology [1, 27, 34]. Confronting the public with realistic face detection and recommendation applications has proven helpful to inform the ethics discussion by way of facilitating public debate [41, 42, 48]. Through their examination of constant surveillance in public space, these studies revealed multidimensional judgments about privacy and the continuous quest to balance social concerns with convenience. In response to increasing ethical challenges, several academic and industry initiatives established frameworks to warrant responsible use of facial analysis, automated decision-making and, in a broader context, artificial intelligence (AI) [5, 15, 74, 84]. Many of these frameworks originate in the deontological philosophy of Kant, establishing fundamental insight into the rightness or wrongness of actions and reflecting accepted standards of behavior [38]. The initiatives coincide with the emergence of conferences, calls for special issues, and creation of standards committees, working groups and codes of ethics [6, 39, 40] that aim to formulate responses to ethical challenges faced by technologists and innovators.

Ethics and Values Research in HCI

Recent HCI initiatives complement these ethics frameworks by engaging with members of the public to better understand beliefs and judgments about AI and its implications [1, 27, 34, 60, 71, 88]. Most of these initiatives involve methods such as surveys, workshops and interviews, where the study aims are fully disclosed upfront. This can be explained by the use of common ethical principles that are core to the HCI community's approach. HCI builds upon a long tradition of establishing and adhering to ethical guidelines that go beyond purely legal requirements [59]. This tradition is exemplified by SIGCHI's establishment of an internal ethics committee [39], and the community's ongoing commitment to set appropriate measures that protect the emotional and mental well-being of researchers and participants [64, 83, 87]. The goal is to enable rich insights while avoiding moral conflict and protecting the psychological state of human participants.

In-situ studies and experience samples are beneficial for gaining insights into the user's perspective, for example, about how the use of wearable technologies may ultimately violate the privacy of non-users [46, 51, 56, 57, 78]. However, unique challenges arise when HCI research takes place in natural contexts, where technologies and experiences are evaluated in-situ and where there is less control over the experiences and behaviors of human participants [19, 22, 23, 72]. Studies in uncontrolled environments must carefully negotiate the personal boundaries of participants [9]. For instance, publicly exposing participant behaviors and interactions may create a sense of embarrassment among them or awkwardness among bystanders [18]. An additional challenge lies in seeking fully informed consent from participants while attempting to study natural responses and providing suitable opportunities for withdrawal and disengagement [10].

Designing for Speculation on Ethical Implications

Public responses to implications of technology have been previously explored by way of design fiction [8, 13, 80]. By envisioning a future scenario, design fiction represents a possible future that may become real and which opens up space for discussion [31, 35, 81]. Presenting themselves only as possibilities, fictional scenarios pose the question "what if?", inviting reflection on the potential impact of technological developments on societal behaviors [30]. The approach relies on envisioning convincing artifacts that balance surprise with feasibility, leaving the spectator to wonder about how far they draw on existing technologies and how much of this is an invention. However, a unique opportunity lies in the evaluation of technology design demonstrators that generate disagreement and challenge dominant practices [29, p. 115]. Their confrontational nature enables the public to participate in debate and express opinions on inherent challenges.

Recent studies embraced speculative qualities to interrogate the public's understanding of ethics and values by way of artistic representation. For instance, *Quantified Self* immersed spectators in a theater performance to provoke speculation about the use of personal data by AI systems [79]. Besides its artistic merit, the study revealed that confrontation with an ethically challenging technology helped people recognize the need to understand the usage and sharing of personal data by companies. Yet, while the value of artistic representations has proven to benefit audience engagement [33, 47], we believe it is equally important to stimulate discussion about ethics by provoking response through design-oriented objects that reflect real-life application scenarios and that exemplify prob-

lematic aspects of an emerging technology. Enabling the public to walk up to a realistic demonstrator that blends in with its environment, to interact with it, and to experience individual implications seems a relevant approach to collect ecologically valid feedback on ethical use of emerging technologies.

BIOMETRIC MIRROR

We developed Biometric Mirror to explore common understandings of the inner workings of facial analysis technology and opinions about the ethical concerns about automated decision-making. To accomplish these goals, we established four design requirements.

Realistic Materiality: Unlike early stage prototyping methods, such as Wizard of Oz [52], we aimed to realistically represent application scenarios of facial analysis technology and algorithmic decision-making consequences. Hence, through its physical embodiment and functional behavior, we chose to reflect characteristics of real-life systems that may have been deployed for commercial or surveillance purposes.

Natural Interaction and Environment: We aimed to encourage opportunistic user interactions and thus to collect naturalistic feedback in a physical context that is relevant to the application environment of facial analysis technology. As evaluations took place 'in the wild' [73], we aimed to support people in their reflection on a realistic application scenario and to allow for spontaneous discussions between users, bystanders, and onlookers to emerge.

Embody Concerns: Recent reports and academic work highlight how facial analysis models are used to infer sensitive data from facial expressions, including perceived traits and human behaviors [32, 36, 49, 86]. We chose to embed these concerns in Biometric Mirror, particularly the lack of transparency about the inner workings of facial analysis models, and the absence of any opportunity for the public to question and critique the validity of inferences.

Well-being: Our study aimed to involve those people that are prone to surveillance in the debate about ethics of automated decision-making. Specifically, we involved them personally in the narrative of a surveillance application in an effort to maximize impact while avoiding harm. We reflected on the ethics of our work throughout the study design and implementation by putting appropriate mechanisms in place to protect users before, during and after their interaction. We ensured that the design and functionality of Biometric Mirror acknowledged users' rights, health and well-being, and enabled individual agency over the experience.

As the research team included people with diverse gender, cultural and racial identities, we collectively sought to balance emerging social and ethical concerns of our study with design, functionality and experience responses. We illustrate the nature of our considerations in the following sections.

Implementation

Based on the design requirements, Biometric Mirror was developed for interaction via large public displays through mid-air gestures in a public University space. With a goal to explore issues surrounding the deployment of facial analysis in public spaces, we decided that a public display was a suitable mechanism to visualize the ubiquity of surveillance and the extent of the decisions that could be made by facial analysis algorithms. We carefully designed the facial analysis model and the gesture interaction metaphors to support this purpose.

Facial Analysis

Biometric Mirror classifies faces based on a facial analysis model built on top of the publicly available *10k US Adult Faces Database*, which contains 10,168 natural face photographs [7]. Of these, 2,222 photographs also contain subjective ratings crowd-sourced from 1,274 respondents on perceived demographic, psychological and social attributes including race, facial attractiveness, aggressiveness, and emotional instability. Each respondent rated an average of 26.24 photos (SD=68.04). Ratings correspond to a value in the [1..10] range with 1 and 10 indicating a low- and high-value rating, respectively.

We built a facial analysis model using Microsoft Azure Face and Custom Vision based on the 2,222 tagged face photographs and trained it to predict psychological and social impressions. Due to the small size of the image set per some of the tagged attributes, we mapped the range of ratings to a smaller scale [1..3] with 1, 2 and 3 indicating a 'low', 'average' and 'high' value judgment respectively. The *10k US Adult Faces Database* came with five pre-categorized races, i.e. Caucasian (82.67% of photos), African (9.95%), Hispanic (3.24%), Asian (3.06%) and Middle Eastern (1.08%). Next to binary gender, age and race, we selected 11 attributes that were considered to provoke a response from people while raising suspicion about the psychological quality of the analysis. We chose to display emotion, kindness, happiness, commonness, responsibility, attractiveness, sociability, introversion, aggressiveness, weirdness, and emotional stability.

Ethical Consideration: We trained the classifier with the labels available in the dataset in order to surface the harmful issues created by them (e.g. binary gender representation rather than self-described or non-categorical gender). This design decision resulted in a high likeliness of misjudgments but symptomises the common treatment of gender as a binary and physiological phenomenon in most research and commercial gender recognition software [53]. The risk of misidentifying personal characteristics –gender and race– led us to develop the possibility for people to walk away from the screen, upon which the session terminated and all information was erased. In this study, the terms ethnicity and race are used interchangeably. We considered ethnicity less volatile terminology and more suitable for our purpose of mimicking a system that analyses a range of visual, sociological, biological and societal characteristics, such as social norm and cultural tradition. For us, this was particularly important given the US-centricity of the dataset's racial categories and our deployment in Australia. Contact details for the University's counseling service were supplied on a printed document placed next to Biometric Mirror and online. Furthermore, while the initial dataset contained values for 'emotional instability', we recognized potential harm in this particular wording and rephrased to 'emotional stability'.

Gesture Interaction

Users interacted with Biometric Mirror through two mid-air gestures captured and processed by a Microsoft Kinect depth sensor. The setup enabled interaction to remain device-free, without the need for approaching the display and, for instance, touching it [26]. We developed an algorithm that enabled the depth sensor to detect two custom gestures. The first enabled *consent*, with on-screen prompts inviting users to raise a hand in order to consent to their participation. The second enabled *withdrawal*, where users were able to place their hands over their eyes in order to terminate the session. The same effect could be achieved by walking away from the screen.

Ethical Consideration: Gestures were required to be morphologically sufficiently different to avoid unintended triggers of undesired actions. We considered it vital that, for instance, covering eyes (terminate interaction) would never be mistaken for raising a hand (consenting to proceed). We extensively tested the performance of gesture recognition algorithms in the lab and refined them during the pilot and field studies to pick up the supposed behavior accurately.

Interface and Interaction Design

The attraction screen of Biometric Mirror needed to trigger interest from passers-by. We designed it to be modern in appearance, featuring bold sans-serif typography and a vibrant color gradient . Its semitransparent background displayed a real-time, black-and-white wide-angle camera feed and mirror image, allowing passers-by to recognize the dynamic nature of the application [66]. The on-screen prompts asked passers-by if they "*want to see what computers know about [them]*" and revealed that proceeding required consenting by raising a hand. Upon recording consent, the wide-angle camera feed was cropped to focus only on the consenting user.

Subsequent screens revealed a low-resolution and non-graphical user interface that resembled a Unix-based application. With a color palette featuring white and blue, the interface was designed to reveal the data access layer that is often concealed behind the presentation layer of modern, high-resolution applications and that is only fully understood by and accessible to developers [3]. Elements were laid out across two vertical halves with a real-time camera feed of the user's face on the left and results of the analysis, requests for consent and other notifications on the right. Users subsequently proceeded through four steps:

1. First, the *Briefing Screen* contains a prompt that articulates the growing popularity of facial analysis applications (see Figure 1(a)). It encourages the user to move closer toward the screen. The user is then asked to consent for a photo to be taken and analysed by the facial analysis model, and for the analysis to appear (see Figure 1(b)). Upon consenting, a single photo is taken and analysed, a session is created, and the unique session number appears on-screen.
2. The user then viewed *Psychometric Data*. We opted for a minimalist 3-column design that only showed the psychometric attribute, the value for the current user's analysis and the algorithm's confidence value. We displayed information in this manner as we perceived it to be the most ethical manner—it reflected the analytic nature of how such processes would typically work (e.g. as a terminal process or mainframe application), and illustrated that no further interpretation had happened in the background. First, the user viewed age and gender, both appearing at 3-second intervals. The session was then interrupted and the user was asked to consent before continuing. Next, all 12 remaining attributes, values and confidence values were added to the list, appearing at 3-second intervals (see Figure 1(c)).
3. The subsequent *Scenario Screen* displayed a personalized, speculative consequence of the analysis in an algorithmic decision-making process (see Figure 1(d)). One scenario was selected from a list of 12, based on the psychometric attribute with the highest confidence value. For instance, if a user was perceived to be aggressive, the scenario questioned how the user feels about "*all data being shared with law enforcement, allowing them to monitor every movement*". Other scenarios referenced decisions about employment decisions (e.g. high introversion suggested "*data being shared with recruitment agencies, and they exclude you for all management positions*") and health care (e.g. high weirdness suggested "*data being shared with health professionals, urging them to offer you counseling*").
4. Ultimately, upon covering eyes or walking away, a *Debrief Screen* was shown (see Figure 1(e)). The text encouraged public debate, with a subsequent screen providing instructions to remove session data (see Figure 1(f)). The screen appeared as soon as the sensor detected a withdrawal gesture, regardless of users' progress within the application.

Ethical Consideration: Though ambiguous to mimic as much as possible a real-life surveillance application, the briefing screen offered users a general feel of what to expect before consenting to participate more fully. We recognized that misjudged gender, age or race could cause harm, such as previously illustrated in research on harms inflicted upon transgender people [50], and we intended to offer participants sufficient opportunity to withdraw. To inform users in more detail about the study, we displayed the URL of an online project page on each individual screen of Biometric Mirror. The page contained descriptions about the study context, an FAQ, explanations about each individual data attribute, and a list of key contacts including counseling services and the responsible researcher. The project page served as the first point of reference in case any concern would be raised. Scenarios where speculatively phrased as "imagine that..." to prevent users from thinking that decisions were actually carried out.

Pilot Lab Study

Prior to public deployment, we conducted a two-week lab study to assess initial response to Biometric Mirror and to evaluate the success of the interaction modality. While the lab study only enabled research staff to participate, the setup resembled the final field study setup in many ways, including its opportunistic interaction abilities, the absence of on-site research staff, and the availability of a Plain Language Statement (PLS) to briefly describe the purpose and sources for more information. The PLS also contained instructions for users to share concerns with the research team.

Figure 1. User interface of Biometric Mirror through consecutive stages, including (a) briefing; (b) consent; (c) psychometrics; (d) scenario; (e) debrief; and (f) opt-out procedure.

During the study, the responsible researcher was made aware of two researchers' concerns about the ethics of the study. Through conversation, these concerns were attributed to the unavailability of a public information resource that provided more context about the study, such as its justification, procedure, aim, and explanation about the data.

Risk Assessment and Communication Protocol

The psychological disruptiveness of misjudging people's social identity is widely recognized, in particular as it undermines their social status and results in negative affect [14, 62]. This occurs in terms of misjudging gender [50], race and ethnicity [21], and personality and psychological state [25]. Our decision to train the classifier with labels as they were provided in the original dataset, resulted in a high likeliness of misjudgments. In fact, most predictions turned out to be close to random, i.e. precision=58.7% and recall=53.3% based on k-fold cross validation. Hence, we felt that our study posed two real risks. First and foremost, we recognized that our study had the potential to cause harm among its users, including among vulnerable populations that may already have been affected by misjudgments. A second risk was the possible misperception that our study aimed to reinforce norms that may seem to make the oppression of vulnerable groups acceptable. However, the goal of the study was specifically to explore

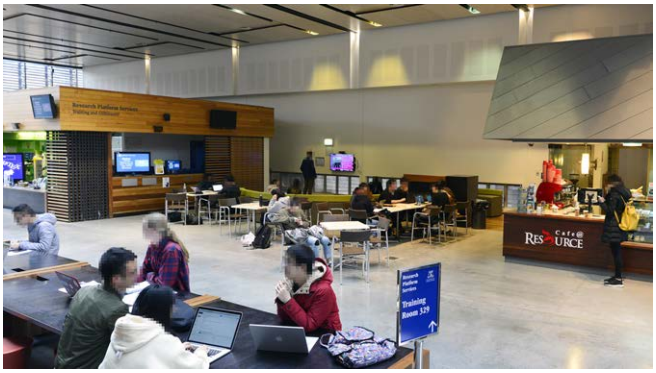

Figure 2. Biometric Mirror (middle of photo) is set up in a public space on a University campus. The location is adjacent to a library, cafe and research support office which attracts students, academic staff and members of the public throughout the day.

response to and raise awareness of (mis)judgments and decisions made routinely by automated systems and without the appropriate controls.

The risk assessment concluded that, in addition to our ethical considerations, a proactive communication strategy would be helpful in mitigating harms. We devised a strategy to generate public interest and to provide an information resource for users and the public to learn about the rationale behind this research. This involved writing an article for two local mainstream media outlets about Biometric Mirror and its purpose set to be published on the study launch day [89]. In line with common research practice, we obtained ethical approval to conduct our study from the Human Research Ethics Committee of The University of Melbourne. The ethics application reflected the purpose and setup of our communication strategy. Contacts within the Chancellor's Office and the Faculty's media team issued a range of 'holding statements' to key media representatives within the University. The approach is common in corporate communication strategy [12] and involved responses to anticipated questions that could be shared in case of critique. This illustrates that the research team and the University both understood the value of the research while acknowledging the challenge of adequately responding to concerns and critique.

IN-THE-WILD STUDY

Biometric Mirror was installed on an interactive display in a public space on the campus of the University of Melbourne for 50 consecutive days (see Figure 2). The space is located adjacent to one of the University's main libraries and is publicly accessible from 8am to 9pm daily. The library supports the faculties of science, engineering, and arts, and thus attracts a diverse visitor profile. The atmosphere in the space is calm with a diverse audience consisting primarily of students, academic staff and members of the public visiting the library.

Method

Three researchers observed the behaviors taking place around Biometric Mirror and conducted semi-structured interviews with users during the first 21 days of deployment. Questions aimed to elicit responses regarding the perceived accuracy of the analysis and about the ethical implications of Biometric Mirror and other facial analysis applications they were aware of. We asked users specifically about their thoughts on the transparent and speculative nature of Biometric Mirror. Interviews were audio recorded and transcribed, and additional field notes were taken on-site. Notes, interviews, and observations from the three researchers were combined and reviewed collaboratively to develop a shared interpretation of public attitudes towards Biometric Mirror, facial analysis applications, and the broader theme of digital ethics.

We captured all interaction and facial analysis data from users of Biometric Mirror in timestamped data logs. During active sessions, we used the Kinect's built-in body tracking functionality to log the number of bystanders. No other analysis was performed on the bodies or faces of bystanders, since they did not necessarily consent to participating in the study. All logged data was later segmented into tagged information to identify the total number of users, number of bystanders during an interaction session, data points at which interaction was abandoned, and number of users who had requested for their data to be removed. In addition, we kept track of media inquiries and publications that appeared during the 50 days of deployment. We observed discussions on social media that featured the *#BiometricMirror* hashtag in order to keep track of and understand the general nature of public debate.

Results

A total of 798 interactions took place with Biometric Mirror over 50 days. 400 users were identified by Biometric Mirror to be female, 398 to be male. 653 users (81.70%) completed the full interaction sequence, with the attributes 'age' and 'attractiveness' being the main drop-off points along the way (respectively n=77, 9.64% and n=13, 1.62%). Users interacted with Biometric Mirror in almost equal parts individually (n=267, 33.46%), in groups of two (n=270, 33.83%) and in groups of three or more (n=261, 32.71%). Interaction data excludes one request that we received from a user for session data to be eliminated from the study. Three researchers spent a total of 51 hours on-site in the first 21 days after deployment to observe interactions with Biometric Mirror and recruit interview participants. We interviewed a total of 40 people from various ethnic origins after they concluded their interaction. 17 interviewees self-identified as female and 23 as male. Interviews took approximately 6 minutes each.

During the 50 days of deployment, 155 news articles were published in online media outlets across 20 countries, reaching close to 204 million readers according to our media metrics provider. In addition, 2,588 messages were published on social media responding to the study or featuring the *#BiometricMirror* hashtag. None of the holding statements needed to be issued. Besides requests from the media to engage with the technology, several emails were sent to the research team from members of the public that hoped to interact with Biometric Mirror via an online website or mobile app. Four other requests originated from industry. This included an overseas face recognition technology supplier that sought to incorporate Biometric Mirror in their product offering and representatives from two local and one overseas recruitment companies to discuss collaboration opportunities.

FINDINGS AND DISCUSSION

One of the most salient observations made during the study was that no single user voiced concern about the impact of the study via any of the channels advertised on the PLS. We observed that many users perceived the psychometric analysis as a leisurely activity, sparking enjoyment as the analysis appeared on the screen, regardless of the often harsh personal analyses. Biometric Mirror provoked users to speculate about the inner workings of the system, the consequences that such a system might present, and the extent to which facial analysis and automated decision-making are increasingly present in society. In this section, we investigate public opinions about the ethics of facial analysis applications.

Awareness About Common Application Scenarios

We learned that a large majority of users seemingly underestimated the extent of common use cases for facial analysis and automated decision-making technology. Typically, when asked about scenarios they were aware of, these were only imagined to be in the context of security, such as for crowd surveillance, law enforcement, and forensic investigations, i.e. policing scenarios: "*Right now it's probably mostly facial recognition and just being able to find people [...] and looking up passport data*" (P01). Most users felt comfortable with such applications because of their contribution to public safety. Yet, the specific output of Biometric Mirror's speculative scenario made participants reflect on the potential to be the victim of discrimination ("*I [would] be concerned if an immigration officer thinks I'm aggressive*", P15).

Only when we informed participants at the end of the interview that similar technology is also used in retail environments, for recruitment purposes and is increasingly governing society in Chinese cities, they expressed concern about the possibility to 'profile' people. They noted the challenges that come with the responsible processing and handling of facial analysis data. In fact, data permanence was a significant concern and users reflected on the implications of having a flawed readout stored forever ("*it should not be held on to forever, because it is often just a snapshot*", P07) or used against you ("*imagine such information is used in the judicial system*", P09), and the mechanisms that are put in place to the warrant integrity of data and people ("*Privacy should be a major concern. It's important to make sure the data is not breached*", P05).

The lack of awareness around the presence of facial analysis technologies in public space concerned users: "*I'm okay with law enforcement getting information, but the general public should know that they are doing this*" (P14). Biometric Mirror made users reflect on the potential for functionality and inner workings of algorithms to be made visible. They argued for increased transparency around the presence and use of facial analysis technology ("*I'd like to know what more was behind getting all those, the analysis for myself*", P20). Here, users identified a need for more public awareness about the mechanisms that inform facial analysis processes: "*there [needs to] be somebody who can provide for any users an understanding of the process, what's going on, that would be useful.*" (P19).

Our interviews suggest that the public still underestimates scenarios where facial analysis and algorithmic decision-making are currently in use. While this is surprising because of increasing mainstream media coverage about the technology, it is not entirely unexpected. Most coverage seems to focus on specific and high-profile applications, such as automated CCTV analyses and the so-called Chinese Social Credit System, and refrains from highlighting the insidious nature of unconsented applications in other contexts that are often much closer to home, such as recruiting, retail and workplace monitoring. Addressing users by way of speculative scenarios, made them realize broader societal impact of automated decision-making applications. Regardless of users' agreement with surveillance tools for political, military, social or economic reasons, the scenario illustrated unequivocally how each individual can, at any point, be affected and harmed by automated decision-making applications without control over its validity or accuracy.

Fallibility of Psychometric Profiling

We observed that the system prompted curiosity and speculation about the underlying technology, and elicited reflection on concerning application scenarios: "*I can imagine [this] can take a photo of you, do a sentiment analysis and tweet that or whatever. And then cops show up out of nowhere and get you*" (P08). Several participants assumed that the internal logic of Biometric Mirror was 'flawed' (P01, P11, P39), particularly in response to their information being clearly misidentified. Belief in the system was most often lost when Biometric Mirror clearly mischaracterized a user's race. This was not entirely uncommon, given the significant majority of images in the dataset tagged as Caucasian.

Users, social media conversations, and press regularly drew parallels between Biometric Mirror and horoscopes or phrenology, acknowledging the danger of using such information for automated decision-making purposes. In fact, most participants seemed not to put faith in the system's psychological analysis, often due to a perceived mismatch between the readouts and their self-perception, arguing that "*It's cool to see how it interprets us, but then it's also interesting to see [that] it doesn't reflect the way we see ourselves.*" (P18). Here, despite the apparent ludic quality, participants questioned the use of facial analysis on the basis of self-perception. A user who ranked low for responsibility expressed his disbelief in the system because "*I've been a carer to my disabled brother*" (P19). These findings suggest that the direct interaction with a demonstrator such as Biometric Mirror made participants recognize limitations of facial analysis for psychometric profiling, such as due to its inability to thoroughly assess a person's psychological state. However, it did not prevent them from envisaging problematic applications and use cases: "*Imagine a system that incorrectly identifies qualities someone doesn't have. They would be disadvantaged because they're at the mercy of AI*" (P04).

While Biometric Mirror was unambiguous in illustrating its core functionality of analyzing faces, it remained silent on explaining its internal logic. The system represented a 'black box' [76] that provided few clues about its operations. Using the system allowed participants to speculate about its inner workings: "*It's the little ways that your face is moving in. Like, if you're happy your eyes might be a bit more squinty or lips*

will curve up." (P07). Unlike most others, several participants stressed their belief in the validity of the assumptions made by the system: "*[The readout] must be right. Because the assessment is made by a computer, and computers are better than people at drawing such conclusions.*" (P03). In fact, the readout seemed to even make some participants question their understanding of themselves: "*I'm an introvert in some parts, but... I don't think I'm an introvert, you know... No. I don't think I'm an introvert at all. I mean, I am, it depends*" (P07).

Our study encouraged users to rethink legitimacy and fairness of analyses based on camera footage. We observed a misunderstanding about the objectiveness of algorithms that intrinsically rely on subjective (i.e. crowd-sourced) data. This opens up a compelling space for public awareness campaigns that illustrate in accessible, engaging and understandable ways the conceptual workings of processes that underpin facial analysis technology, such as crowd-sourcing and machine learning.

Consequences of Algorithmic Decision-Making

One of our initial concerns in deploying Biometric Mirror was that participants might respond to the psychometric readout with disgust or concern, such as after being rated with 'low' attractiveness or 'high' aggressiveness. However, in observing participants' interactions and appraisal of the system, we found that they largely responded in a way that was ludic and playful. One person claimed that using Biometric Mirror was "*definitely fun. Especially for groups of people like friends*" (P16). We observed participants laughing in response to both their own readouts and to those of their peers, and that participants would often take a photograph of their analysis for sharing and comparing with other people. Rather than an insidious or sinister endeavor, people saw the study as a "*social experiment*" (P06, P07, P25, P30) and assigned a degree of agency to the system ("*It was a computer gone rogue, insulting people as if was having a bad day.*", P31), similar to how people attribute human characteristics to computers upon seeing unexpected or undesired results [68].

However, the playful behaviors did not prevent participants from imagining possible futures and speculating about the broader consequences of ubiquitous facial analysis—particularly in scenarios where data is unreliable or flawed ("*Imagine a world where you have a 'suspicious' face. It would make you worried about buying a pressure cooker. It would change the way you act*", P02). In considering the fairness of facial analyses, participants recognized the extent to which Biometric Mirror embodied two key ethical concerns: the use of automated facial analysis to determine subjective attributes, and the potential for this analysis to enable decision-making. They questioned the validity of judging personality based on facial attributes and recognized the potential for such analyses to promote discrimination and bias [88], such as by law enforcement agencies and industry. For instance, some participants recognized application scenarios of facial analysis in human resources (P14, P28). Here, as potential misuse of facial analysis was identified, users called for sensitivity in its usage: "*I don't know how reliably personality and appearance can be correlated like that. It makes sense to use if it is reliable, otherwise it would just encourage prejudice*" (P26). One participant who worked in welfare recognized that automated analysis should not be used in such settings: "*there's always the danger of falling back on profiling. You've got to check yourself against [it]. With computers there's not much room for subtleties that you get when you're speaking to someone.*" (P39). He argued that some environments were simply too 'sensitive' to leave any form of decision-making up to algorithms and computers: "*It's either this or that with a computer. There's no room for subtleties that you would get when speaking to someone*" (P39).

Our decision to conclude interactions with a personalized, speculative scenario proved helpful for people to reflect upon the implications of endemic facial analysis for society at large ("*If you look at the way [technology] is being used, there are some alarming political implications*", P13), referring to some global cases that received significant media coverage in recent times [4, 16]. Others speculated about decisions based on the analysis ("*imagine if you're offered a job, and it says oh you're too aggressive. That would really suck. That's a disturbing future, right there.*", P24). Some users realized that, despite the growing use of algorithmic decision-making processes, the technology could easily produce unreliable results with significant consequences.

Transparency, Consent and Well-Being

The deployment of Biometric Mirror into a public space supported participants in reflecting about the extent to which surveillance is used in everyday public settings. Participants recognized the value of deploying Biometric Mirror in a University space that is considered "*safe*" (P01, P31, P14) as studies are bound by research ethics. As the space contained some permanent surveillance cameras, some participants questioned their inability to identify "*what [these cameras] capture and what the information is being used for*" (P32). While not so much a concern for applications in a University space, some participants questioned such technology being used for commercial benefit such as in shopping malls. It made them reflect on solutions for greater transparency about the data that some surveillance systems may be able to retrieve from facial analyses: "*An air of transparency would probably go a long way*" (P19). Participants suggested the need for solutions, consisting of labels at doors and next to cameras that serve as a privacy policy ("*They should outline what data is captured, how it is processed and why*", P11), and opportunities for the public to review and amend the data captured by analysis technologies, such as by way of mobile applications (P11). Similar recommendations were made in the context of research on gender recognition, encouraging the support of self-expression and autonomy by way of defining and modifying one's own gender identity [43].

Consenting to interaction with Biometric Mirror occurred via a relatively harmless mid-air gesture. It made users recognize the complexity of refusing consent to camera surveillance: "*I suppose it is unethical — well, I guess [Biometric Mirror] is ethical because it asks for your permission, but when you're just out in public [space], you don't really consent to being surveilled*" (P40). Most users realized that "*the only way to opt-out is to not enter [a] shopping mall*" (P11) which

in practice may be undesirable or even impossible given the ubiquity of the technology and the convenience of shopping malls. While most users only referenced the challenges of consenting to surveillance and facial analysis systems, some also questioned the privacy implications of consenting to use other technologies, such as how public WiFi networks in retail environments keep track of users' location data to inform business processes [82] or how Augmented Reality glasses record faces without prior consent [28].

The decision to require consent multiple times throughout the interaction was considered "*good, because it does ask for consent each step of the way*" (P18) and made users feel "*safe, as this is a study bound by research ethics*" (P01). Yet, as we reviewed interaction data, we noticed that on one occasion a bystander performed the gestures on behalf of a user. Here, the bystander put her hand in front of the active user's face, signaling a desire to withdraw from interaction. While this occurrence is certainly harmless (the session ended abruptly), it made us reflect that the opposite could potentially have happened as well: in theory, a close bystander could raise a hand to consent on behalf of another user, thereby signaling Biometric Mirror to continue displaying data and potentially upsetting the user. The activity, which we call *interaction hijacking*, may seem harmless at first but requires particular attention in the context of cases that confront users with ethical implications, since a hijacked gesture or behavior may set in motion unintended consequences that affect users' well-being.

Separate reflections on the consent mechanism emerged on social networks. Shortly after Biometric Mirror launched and press coverage appeared, we observed a Twitter thread consisting of 25 scenarios that reflected on our withdrawal gesture by covering the eyes [45]. The user speculated about a range of fictional scenarios that featured the covering eyes gesture of Biometric Mirror as a standard for revoking consent, renamed as 'wiping'. Some scenarios reflected ongoing discussions on the ethics of facial analysis research ("*[A] research group is met with widespread outrage when they claim that wipe style is correlated with homosexuality after analyzing anonymous wipes from cameras outside gay bars and strip clubs.*"), while others mirror our response from users requesting opportunities to revoke consent ("*People start to wipe as a habit when entering a shop or turning a street corner.*"). The spontaneous emergence of the thread and its thought-provoking scenarios illustrate the quality of demonstrators such as Biometric Mirror to stimulate creative thinking about privacy mechanisms beyond the physical space where it is deployed.

These findings suggest that Biometric Mirror has elicited reflection on a complex and potentially harmful ethical question in a manner that is provocative and playful while remaining true to the goal of ethical research. There seems to be broad public interest in having access to more consent, opt-in and opt-out procedures for applications that deal with sensitive, personal data. Our findings indicate an opportunity for more explainability and dialogue to be integrated in analysis and decision-making applications by design.

Media, Public Requests and Industry Engagement

Our decision to proactively communicate about the study by way of two articles in general science publications proved useful in several ways. First, it set the right tone about the research goal, thereby minimizing the risk of public backlash and a general misunderstanding of our objectives. Second, the publications triggered an organic uptake by other media and social media users to engage in a conversation about the ethics of some current technologies. And third, it helped to attract people to visit the space where Biometric Mirror was set up. Often, this included people from outside the University community, citing their appreciation for the study set-up and the opportunity to better understand potential personal implications of automated decision-making (P06, P09, P11).

The emails that we received from members of the public typically inquired about how to experience Biometric Mirror for themselves without having to pay a visit to the University campus. By contrast, the requests from industry displayed an interest in making use of Biometric Mirror in several ways. We corresponded via email with one of the local recruitment companies that contacted us, and we learned that there was interest to incorporate our psychometric analysis algorithm in existing recruitment processes. This reflects practices that are already occurring in the recruitment industry, such as those used by video recruitment providers that parse video footage through personality engines to detect a range of behavioral metrics for the job applicant [24]. Given the unreliability of our algorithm, it is needless to say that we did not further proceed with these initiatives. The differences in media reporting can explain some of the confusion about the suitability of Biometric Mirror for business processes. We found it particularly telling that while most media correctly reported about our algorithm in the context of addressing pressing ethical concerns, some overseas media failed to report on the ethical objectives. Instead, their reports only mentioned that we had developed an algorithm to distinguish personality traits from a single photograph. While the study has currently ended, media and speaking opportunities are continuing. To date, this includes a total of 27 local and global opportunities to present Biometric Mirror and to make the discussion on ethics of emerging technologies more tangible for key stakeholders in the public, political, academic and corporate realm.

Limitations

Out of ethical considerations, we compromised to conduct our study outside a controlled lab environment but in a public space on a University campus. This enabled us to acquire permission, conveniently observe and interview participants, and immediately intervene if needed. The public nature of Biometric Mirror elicited unique feedback that we believe would otherwise not have been revealed in lab studies. However, despite being in full control over the study and working within a sound ethical framework, we are convinced that more stark response would have been collected in other public environments, such as city centers and transportation hubs. Yet, in our setup there is still a cause for concern. We did not further articulate the potential consequences of our study besides our collective ethical considerations as captured in the study design. Furthermore, researchers were unable to attend the setup

continuously and thus unable to follow up with each participant to assess whether they felt harmed. While we believe the risk of harm was no greater than that encountered via interaction with currently existing systems, we realize that we may not have sufficiently understood the risk of drawing attention to misrepresentations, especially for vulnerable populations. This raises important questions for this kind of research, such as, what if users struggle with their analysis at a later time and have no access to the necessary project information? What happens when people are put under pressure by friends or strangers that observed their analysis? And more importantly, where should we draw the line between provocation for the sake of research and the risk for causing harm? Future work may unpack these questions and investigate how informed consent and debrief procedures can mitigate risks.

We had anticipated response from users with regards to gender representation, but failed to collect any. This observation does not mean there is currently no debate or concern about binary gender representation [43, 75, 53]. Instead, we believe that while our approach was successful in attracting spontaneous users and stimulating public discussion, it also introduced the challenge to reach a heterogeneous sample of the population. The resemblance of a real-life, harmful application—vital to stimulate discussion—may actually have prevented those that often fall victim to surveillance and profiling from participating and sharing opinions and concerns.

DESIGN OPPORTUNITIES

Based on our findings, we share the following pointers for future studies that aim to develop interactive demonstrators to interrogate the understanding of pressing ethical challenges:

- The public needs an opportunity to uncover the issues that the demonstrator highlights, such as by recognizing the personal implications of an emerging technology and its long-term effects on society. Here, we suggest future work to give users a personalized view of the issues at hand, such as by integrating narratives within the user experience or presenting an outlook onto a speculative future.
- Future work must spark public debate about ethical challenges, and thus are best deployed in the field. Researchers must consider the personal and social behaviors reflective of the proposed study location. However, as ethics research is now conducted in an uncontrolled environment, researchers must assess adverse impact of the study on users. The study goals should balance the functional and design considerations to enable a morally sound presentation and experience, as well as consent, briefing and debriefing mechanisms.
- In their design and functionality, future demonstrators benefit from a provocative character in order to encourage public debate. There remains a need to ensure that the research remains ethical, given that negative experiences can harm participants, damage the integrity of the research, and tarnish the reputation of the researchers involved. Hence, we suggest that future endeavors carefully balance the provocative nature of the demonstrator's design and functional characteristics with steps to warrant the well-being of users in a natural environment and their understanding of the technology under investigation.
- While speculative demonstrators of a potential future facilitate debate in their own right, it is vital that the provocative nature is counterbalanced by the availability of information about their purpose and easily accessible communication with the researchers that are involved.

CONCLUSION

An increasing number of emerging technologies affect the fabric of society as reports of unanticipated and unintended usage emerge. The question often remains how well the public is aware of their potential uses and consequences, such as when facial analysis applications use biased data to inform decision-making processes. In this paper, we studied Biometric Mirror, an interactive facial analysis application that presented users with a personalized, speculative scenario of automated decision-making. It demonstrated the potential for a realistic design-oriented object to elicit insightful responses about the ethical implications of facial analysis technology and automated decision-making, with public debate extending well beyond the community that came in direct contact with Biometric Mirror, including social media and press.

We found that users interpreted Biometric Mirror as a ludic artifact that was capable of provoking reflection on the underlying concerns that are associated with facial analysis technology and automated decision-making. Our analysis illustrates that the public has limited awareness about environments where facial analysis applications are implemented and there is a common misunderstanding about the objectiveness and validity of facial analysis algorithms. However, Biometric Mirror made the issues more transparent and supported the public to be involved in a discussion that is otherwise considered technically, socially, politically or culturally complicated. Through their confrontation with a speculative scenario in Biometric Mirror, people recognized the significant challenges that the technology introduces, hence eliciting reflection on opportunities to become aware of when, where, how and why sensitive data is being collected (and how to revoke consent), and how it is being processed to automate decisions that may have far stretching consequences.

Facial analysis, automated decision-making, artificial intelligence and various other new and emerging technologies are predicted to have a significant impact on our daily lives. But they also introduce significant ethical challenges. As such, we believe that there is a particular opportunity for further initiatives that involve members of the public in the debate. This must help the technology to move in a direction that benefits society, rather than entrench and amplify current challenges.

ACKNOWLEDGEMENTS

The authors wish to thank the anonymous reviewers for their valuable insights and contributions, as well as the media advisors at The University of Melbourne and the entire Science Gallery Melbourne team for their ongoing support. Eduardo Velloso is the recipient of an Australian Research Council Discovery Early Career Award (Project Number: DE180100315) funded by the Australian Government.

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2018 (CHI '18)*. ACM, New York, NY, USA, 1–18. DOI: http://dx.doi.org/10.1145/3173574.3174156

[2] Oscar Alvarado and Annika Waern. 2018. Towards Algorithmic Experience: Initial Efforts for Social Media Contexts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2018 (CHI '18)*. ACM, New York, NY, USA, 286–12. DOI: http://dx.doi.org/10.1145/3173574.3173860

[3] Vasilios Andrikopoulos, Tobias Binz, Frank Leymann, and Steve Strauch. 2012. How to Adapt Applications for the Cloud Environment: Challenges and Solutions in Migrating Applications to the Cloud. *Computing* 95, 6 (Dec. 2012), 493–535. DOI: http://dx.doi.org/10.1007/s00607-012-0248-2

[4] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. (2016). Accessed: 2018-12-03.

[5] Thomas Arnold and Matthias Scheutz. 2018. The "Big Red Button" is too Late: An Alternative Model for the Ethical Evaluation of AI Systems. *Ethics and Information Technology* 20, 1 (Jan. 2018), 59–69. DOI: http://dx.doi.org/10.1007/s10676-018-9447-7

[6] IEEE Standards Association. 2018. Ethically Aligned Design, Version 2 (EADv2). https://ethicsinaction.ieee.org. (2018). Accessed: 2018-12-02.

[7] Wilma A Bainbridge, Phillip Isola, and Aude Oliva. 2013. The Intrinsic Memorability of Face Photographs. *Journal of Experimental Psychology: General* 142, 4 (2013), 1323–1334. DOI: http://dx.doi.org/10.1037/a0033872

[8] Jeffrey Bardzell, Shaowen Bardzell, and Erik Stolterman. 2014. Reading Critical Designs: Supporting Reasoned Interpretations of Critical Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2014 (CHI '14)*. ACM, New York, NY, USA, 1951–1960. DOI: http://dx.doi.org/10.1145/2556288.2557137

[9] Steve Benford, Matt Adams, Ju Row Farr, Nick Tandavanitj, Kirsty Jennings, Chris Greenhalgh, Bob Anderson, Rachel Jacobs, Mike Golembewski, Marina Jirotka, Bernd Carsten Stahl, Job Timmermans, and Gabriella Giannachi. 2015. The Ethical Implications of HCI's Turn to the Cultural. *ACM Transactions on Computer-Human Interaction* 22, 5 (2015), 1–37. DOI: http://dx.doi.org/10.1145/2775107

[10] Steve Benford, Chris Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, and Tom Rodden. 2012. Uncomfortable Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2012 (CHI '12)*. ACM, New York, NY, USA, 2005–2014. DOI: http://dx.doi.org/10.1145/2207676.2208347

[11] Hal Berghel. 2018. Malice Domestic: The Cambridge Analytica Dystopia. *Computer* 51, 5 (2018), 84–89. DOI:http://dx.doi.org/10.1109/MC.2018.2381135

[12] Michael Bland. 2016. *Communicating Out of a Crisis*. Springer, New York, NY, USA.

[13] Mark Blythe, Kristina Andersen, Rachel Clarke, and Peter Wright. 2016. Anti-Solutionist Strategies: Seriously Silly Design Fiction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2016 (CHI '16)*. ACM, New York, NY, USA, 4968–4978. DOI: http://dx.doi.org/10.1145/2858036.2858482

[14] Jennifer K Bosson, Jennifer L Prewitt-Freilino, and Jenel N Taylor. 2005. Role Rigidity: A Problem of Identity Misclassification? *Journal of Personality and Social Psychology* 89, 4 (2005), 552–565. DOI: http://dx.doi.org/10.1037/0022-3514.89.4.552

[15] Nick Bostrom and Eliezer Yudkowsky. 2014. The Ethics of Artificial Intelligence. In *The Cambridge Handbook of Artificial Intelligence*, Keith Frankish and William M Ramsey (Eds.). Cambridge University Press, Cambridge, UK, 316–334. DOI: http://dx.doi.org/10.1017/CBO9781139046855.020

[16] Rachel Botsman. 2017. Big Data Meets Big Brother as China Moves to Rate its Citizens. https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion. (2017). Accessed: 2018-12-21.

[17] Danah Boyd and Kate Crawford. 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological and Scholarly Phenomenon. *Information, Communication & Society* 15, 5 (2012), 662–679. DOI: http://dx.doi.org/10.1080/1369118X.2012.678878

[18] Harry Brignull and Yvonne Rogers. 2003. Enticing People to Interact with Large Public Displays in Public Spaces. In *IFIP TC13 International Conference on Human-Computer Interaction 2003 (Interact '03)*. IOS Press, Zurich, Switserland, 17–24.

[19] Barry Brown, Stuart Reeves, and Scott Sherwood. 2011. Into the Wild: Challenges and Opportunities for Field Trial Methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2011 (CHI '11)*. ACM, New York, NY, USA, 1657–1666. DOI: http://dx.doi.org/10.1145/1978942.1979185

[20] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (FAT '18)*, Sorelle A Friedler and Christo Wilson (Eds.). PMLR, New York, NY, USA, 77–91.

[21] Byron Burkhalter. 1999. Reading Race Online. In *Communities in Cyberspace*, Mark A Smith and Peter Kollock (Eds.). Routledge, New York, NY, USA, 60–75.

[22] John Carroll and Mary Beth Rosson. 2013. Wild at Home: The Neighborhood as a Living Laboratory for HCI. *ACM Trans. Comput.-Hum. Interact.* 20, 3 (2013), 16. DOI:http://dx.doi.org/10.1145/2491500.2491504

[23] Alan Chamberlain, Andy Crabtree, Tom Rodden, Matt Jones, and Yvonne Rogers. 2012. Research in the wild: understanding 'in the wild' approaches to design and development. In *Proceedings of the Conference on Designing Interactive Systems 2012 (DIS '12)*. ACM, New York, NY, USA, 795–796. DOI: http://dx.doi.org/10.1145/2317956.2318078

[24] Tomas Chamorro-Premuzic, Dave Winsborough, Ryne A Sherman, and Robert Hogan. 2016. New Talent Signals: Shiny New Objects or a Brave New World? *Industrial and Organizational Psychology* 9, 03 (May 2016), 621–640. DOI: http://dx.doi.org/10.1017/iop.2016.6

[25] Neil D Christiansen, Shaina Wolcott-Burnam, Jay E Janovics, Gary N Burns, and Stuart W Quirk. 2009. The Good Judge Revisited: Individual Differences in the Accuracy of Personality Judgments. *Human Performance* 18, 2 (Nov. 2009), 123–149. DOI: http://dx.doi.org/10.1207/s15327043hup1802_2

[26] Adrian Clark, Andreas Dünser, Mark Billinghurst, Thammathip Piumsomboon, and David Altimira. 2011. Seamless Interaction in Space. In *Proceedings of the Australian Human-Computer Interaction Conference 2011 (OZCHI '11)*. ACM, New York, NY, USA, 88–97. DOI:http://dx.doi.org/10.1145/2071536.2071549

[27] Anupam Datta, Shayak Sen, and Yair Zick. 2016. Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems. In *Proceedings of the IEEE Symposium on Security and Privacy (SP '16)*. IEEE, San Jose, CA, USA, 598–617. DOI: http://dx.doi.org/10.1109/SP.2016.42

[28] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In Situ with Bystanders of Augmented Reality Glasses: Perspectives on Recording and Privacy-mediating Technologies. In *SIGCHI Conference on Human Factors in Computing Systems 2014*. ACM, New York, New York, USA, 2377–2386. DOI: http://dx.doi.org/10.1145/2556288.2557352

[29] Carl DiSalvo. 2012. *Adversarial Design*. The MIT Press, Cambridge, MA, USA.

[30] Anthony Dunne. 1999. *Hertzian Tales*. Royal College of Art (RCA), London, UK.

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

[32] Natalie C Ebner, Michaela Riediger, and Ulman Lindenberger. 2010. FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods* 42, 1 (2010), 351–362. DOI: http://dx.doi.org/10.3758/BRM.42.1.351

[33] Ernest A Edmonds. 2014. Human Computer Interaction, Art and Experience. In *Interactive Experience in the Digital Age*. Springer International Publishing, Cham, Switzerland, 11–23. DOI: http://dx.doi.org/10.1007/978-3-319-04510-8_2

[34] Michael D Ekstrand, Rezvan Joshaghani, and Hoda Mehrpouyan. 2018. Privacy for All: Ensuring Fair and Equitable Privacy Protections. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. PMLR, New York, NY, USA, 35–47.

[35] Chris Elsden, David Chatting, Abigail C Durrant, Andrew Garbett, Bettina Nissen, John Vines, and David S Kirk. 2017. On Speculative Enactments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 5386–5399. DOI: http://dx.doi.org/10.1145/3025453.3025503

[36] David J Fleet, Michael J Black, Yaser Yacoob, and Allan D Jepson. 2000. Design and Use of Linear Models for Image Motion Analysis. *International Journal of Computer Vision* 36, 3 (2000), 171–193. DOI: http://dx.doi.org/10.1023/A:1008156202475

[37] Luciano Floridi. 2018. Artificial Intelligence, Deepfakes and a Future of Ectypes. *Philosophy & Technology* 31, 3 (2018), 317–321. DOI: http://dx.doi.org/10.1007/s13347-018-0325-3

[38] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines* 28, 4 (Nov. 2018), 689–707. DOI: http://dx.doi.org/10.1007/s11023-018-9482-5

[39] Christopher Frauenberger, Amy S. Bruckman, Cosmin Munteanu, Melissa Densmore, and Jenny Waycott. 2017. Research Ethics in HCI: A Town Hall Meeting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI EA '17)*. ACM, New York, NY, USA, 1295–1299. DOI: http://dx.doi.org/10.1145/3027063.3051135

[40] Sorelle A Friedler and Christo Wilson. 2018. Proceedings of the Conference on Fairness, Accountability, and Transparency: Preface. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (FAT '18)*, Sorelle A Friedler and Christo Wilson (Eds.). PMLR, New York, NY, USA, 1–2. http://proceedings.mlr.press/v81/friedler18a.html

[41] Batya Friedman, Nathan G Freier, Peter H Kahn Jr., Peyina Lin, and Robin Sodeman. 2008. Office Window of the Future? Field-based Analyses of a New Use of a Large Display. *International Journal of Human-Computer Studies* 66, 6 (June 2008), 452–465. DOI:http://dx.doi.org/10.1016/j.ijhcs.2007.12.005

[42] Batya Friedman, Peter H Kahn, Jennifer Hagman, Rachel L Severson, and Brian Gill. 2009. The Watcher and the Watched: Social Judgments about Privacy in a Public Place. In *Media Space 20 + Years of Mediated Life*. Springer, London, London, 145–176. DOI: http://dx.doi.org/10.1007/978-1-84882-483-6_9

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

[44] Dirk Helbing, Bruno S Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen van den Hoven, Roberto V Zicari, and Andrej Zwitter. 2018. Will Democracy Survive Big Data and Artificial Intelligence? In *Towards Digital Enlightenment*. Springer International Publishing, Cham, Switzerland, 73–98. DOI: http://dx.doi.org/10.1007/978-3-319-90869-4_7

[45] Christine Henry. 2018. A Speculative Fiction Thread (Tweet). https://twitter.com/christinelhenry/status/1024345651436507142. (2018). Accessed: 2018-09-14.

[46] Roberto Hoyle, Robert Templeman, Steven Armes, Denise Anthony, David Crandall, and Apu Kapadia. 2014. Privacy Behaviors of Lifeloggers Using Wearable Cameras. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 571–582. DOI: http://dx.doi.org/10.1145/2632048.2632079

[47] Rachel Jacobs, Steve Benford, and Ewa Luger. 2015. Behind The Scenes at HCI's Turn to the Arts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2015 Extended Abstracts (CHI EA'15)*. ACM, New York, NY, USA, 567–578. DOI:http://dx.doi.org/10.1145/2702613.2732513

[48] Gavin Jancke, Gina Danielle Venolia, Jonathan Grudin, J J Cadiz, and Anoop Gupta. 2001. Linking Public Spaces: Technical and Social Issues. In *SIGCHI Conference on Human Factors in Computing Systems 2001 (CHI '01)*. ACM, New York, New York, USA, 530–537. DOI: http://dx.doi.org/10.1145/365024.365352

[49] Takeo Kanade, Jeffrey Cohn, and Yingli Tian. 2000. Comprehensive Database for Facial Expression Analysis. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG '00)*. IEEE Comput. Soc, Grenoble, France, 46–53. DOI:http://dx.doi.org/10.1109/afgr.2000.840611

[50] Stephanie Julia Kapusta. 2016. Misgendering and its Moral Contestability. *Hypatia* 31, 3 (Aug. 2016), 502–519. DOI:http://dx.doi.org/10.1111/hypa.12259

[51] Genovefa Kefalidou, Anya Skatova, Michael Brown, Victoria Shipp, James Pinchin, Paul Kelly, Alan Dix, and Xu Sun. 2014. Enhancing Self-reflection with Wearable Sensors. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services (MobileHCI '14)*. ACM, New York, NY, USA, 577–580. DOI: http://dx.doi.org/10.1145/2628363.2634257

[52] John F. Kelley. 1984. An Iterative Design Methodology for User-friendly Natural Language Office Information Applications. *ACM Trans. Inf. Syst.* 2, 1 (1984), 26–41. DOI:http://dx.doi.org/10.1145/357417.357420

[53] Os Keyes. 2018. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition . In *ACM Conference on Computer-Supported Collaborative Work (CSCW '18)*. ACM, New York, NY, USA, 1–22. DOI: http://dx.doi.org/10.1145/3274357

[54] Hyeongwoo Kim, Pablo Carrido, Ayush Tewari, Weipeng Xu, Justus Thies, Matthias Niessner, Patrick Pérez, Christian Richardt, Michael Zollhöfer, and Christian Theobalt. 2018. Deep Video Portraits. *ACM Trans. Graph.* 37, 4 (2018), 163:1–163:14. DOI: http://dx.doi.org/10.1145/3197517.3201283

[55] Keith Kirkpatrick. 2016. Battling Algorithmic Bias: How Do We Ensure Algorithms Treat Us Fairly? *Commun. ACM* 59, 10 (2016), 16–17. DOI: http://dx.doi.org/10.1145/2983270

[56] Marion Koelle, Katrin Wolf, and Susanne Boll. 2018. Beyond LED Status Lights: Design Requirements of Privacy Notices for Body-worn Cameras. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '18)*. ACM, New York, NY, USA, 177–187. DOI: http://dx.doi.org/10.1145/3173225.3173234

[57] Sze Yin Kwok, Anya Skatova, Victoria Shipp, and Andy Crabtree. 2015. The Ethical Challenges of Experience Sampling Using Wearable Cameras. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI EA '15)*. ACM, New York, NY, USA, 1054–1057. DOI: http://dx.doi.org/10.1145/2786567.2794325

[58] Caitlin Lustig, Katie Pine, Bonnie Nardi, Lilly Irani, Min Kyung Lee, Dawn Nafus, and Christian Sandvig. 2016. Algorithmic Authority: the Ethics, Politics, and Economics of Algorithms that Interpret, Decide, and Manage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2016 (CHI EA '16)*. ACM, New York, NY, USA, 1057–1062. DOI: http://dx.doi.org/10.1145/2851581.2886426

[59] Wendy E Mackay. 1995. Ethics, Lies and Videotape.... In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 1995 (CHI '95)*. ACM, New York, NY, USA, 138–145. DOI: http://dx.doi.org/10.1145/223904.223922

[60] Jennifer Marlow and Jason Wiese. 2017. Surveying User Reactions to Recommendations Based on Inferences Made by Face Detection Technology. In *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17)*. ACM Press, New York, New York, USA, 269–273. DOI: http://dx.doi.org/10.1145/3109859.3109875

[61] Andrew McAfee and Erik Brynjolfsson. 2012. Big Data: The Management Revolution. *Harvard Business Review* 90, 10 (2012), 60–68.

[62] Kevin A McLemore. 2014. Experiences with Misgendering: Identity Misclassification of Transgender Spectrum Individuals. *Self and Identity* 14, 1 (Nov. 2014), 51–74. DOI: http://dx.doi.org/10.1080/15298868.2014.950691

[63] Andrew McStay. 2016. Empathic Media and Advertising: Industry, Policy, Legal and Citizen Perspectives (The Case for Intimacy). *Big Data & Society* 3, 2 (2016), 205395171666686. DOI: http://dx.doi.org/10.1177/2053951716666868

[64] Wendy Moncur. 2013. The Emotional Wellbeing of Researchers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2013 (CHI '13)*. ACM, New York, NY, USA, 1883. DOI: http://dx.doi.org/10.1145/2470654.2466248

[65] Jörg Müller, Juliane Exeler, Markus Buzeck, and Antonio Krüger. 2009. ReflectiveSigns: Digital Signs That Adapt to Audience Attention. In *Proceedings of the Conference on Pervasive Computing and Communications 2009 (Pervasive '09)*. University of Munster, Springer-Verlag, Berlin / Heidelberg, 17–24. DOI:http://dx.doi.org/10.1007/978-3-642-01516-8_3

[66] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. 2012. Looking Glass: a Field Study on Noticing Interactivity of a Shop Window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2012 (CHI '12)*. ACM, New York, NY, USA, 297–306. DOI: http://dx.doi.org/10.1145/2207676.2207718

[67] Jörg Müller, Dennis Wilmsmann, Juliane Exeler, Markus Buzeck, Albrecht Schmidt, Tim Jay, and Antonio Krüger. 2009. Display Blindness: The Effect of Expectations on Attention towards Digital Signage. In *Proceedings of the Conference on Pervasive Computing and Communications 2009*, Hideyuki Tokuda, Michael Beigl, Adrian Friday, A Brush, and Yoshito Tobe (Eds.). Springer Berlin / Heidelberg, Berlin / Heidelberg, 1–8. DOI:http://dx.doi.org/10.1007/978-3-642-01516-8_1

[68] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues* 56, 1 (2000), 81–103. DOI: http://dx.doi.org/10.1111/0022-4537.00153

[69] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, New York, NY, USA. https://nyupress.org/books/9781479837243/

[70] Jahna Otterbacher, Jo Bates, and Paul Clough. 2017. Competent Men and Warm Women: Gender Stereotypes and Backlash in Image Search Results. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2017 (CHI '17)*. ACM, New York, NY, USA, 6620–6631. DOI: http://dx.doi.org/10.1145/3025453.3025727

[71] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2018 (CHI '18)*. ACM, New York, NY, USA, 1–13. DOI:http://dx.doi.org/10.1145/3173574.3173677

[72] Yvonne Rogers. 2012. HCI Theory: Classical, Modern, and Contemporary. *Synthesis Lectures on Human-Centered Informatics* 5, 2 (2012), 1–129. DOI: http://dx.doi.org/10.2200/S00418ED1V01Y201205HCI014

[73] Yvonne Rogers and Paul Marshall. 2017. Research in the Wild. *Synthesis Lectures on Human-Centered Informatics* 10, 3 (2017), i–97. DOI: http://dx.doi.org/10.2200/S00764ED1V01Y201703HCI037

[74] Stuart Russell, Daniel Dewey, and Max Tegmark. 2015. Research Priorities for Robust and Beneficial Artificial Intelligence. *AI Magazine* 36, 4 (Dec. 2015), 105. DOI: http://dx.doi.org/10.1609/aimag.v36i4.2577

[75] Morgan Klaus Scheuerman, Stacy M Branham, and Foad Hamidi. 2018. Safe Spaces and Safe Places: Unpacking Technology-Mediated Experiences of Safety and Harm with Transgender People. In *Proceedings of the Conference on Computer-Supported Collaborative Work (CSCW '18)*. ACM, New York, NY, USA, 1–27. DOI:http://dx.doi.org/10.1145/3274424

[76] Andrew W Senior and Sharathchandra Pankanti. 2011. Privacy Protection and Face Recognition. In *Handbook of Face Recognition*. Springer, London, UK, 671–691. DOI:http://dx.doi.org/10.1007/978-0-85729-932-1_27

[77] Natasha Singer. 2018. Facebook's Push for Facial Recognition Prompts Privacy Alarms. https://www.nytimes.com/2018/07/09/technology/facebook-facial-recognition-privacy.html. (2018). Accessed: 2018-11-18.

[78] Anya Skatova, Victoria E. Shipp, Lee Spacagna, Benjamin Bedwell, Ahmad Beltagui, and Tom Rodden. 2015. Datawear: Self-reflection on the Go or How to Ethically Use Wearable Cameras for Research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI EA '15)*. ACM, New York, NY, USA, 323–326. DOI: http://dx.doi.org/10.1145/2702613.2725450

[79] Michael Warren Skirpan, Jacqueline Cameron, and Tom Yeh. 2018. More Than a Show: Using Personalized Immersive Theater to Educate and Engage the Public in Technology Ethics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2018 (CHI '18)*. ACM, New York, NY, USA, 1–13. DOI: http://dx.doi.org/10.1145/3173574.3174038

[80] Marie Louise Juul Søndergaard and Lone Koefoed Hansen. 2016. PeriodShare: A Bloody Design Fiction. In *Proceedings of the Nordic Conference on Human-Computer Interaction 2016 (NordiCHI '16)*. ACM, New York, NY, USA, 1–6. DOI: http://dx.doi.org/10.1145/2971485.2996748

[81] Miriam Sturdee, Paul Coulton, Joseph G Lindley, Mike Stead, Haider Ali, and Andy Hudson-Smith. 2016. Design Fiction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2016 (CHI '16)*. ACM, New York, NY, USA, 375–386. DOI:http://dx.doi.org/10.1145/2851581.2892574

[82] Karen P Tang, Pedram Keyani, James Fogarty, and Jason I Hong. 2006. Putting People in Their Place: An Anonymous and Privacy-Sensitive Approach to Collecting Sensed Data in Location-Based Applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2006 (CHI '06)*. ACM, New York, NY, USA, 93. DOI: http://dx.doi.org/10.1145/1124772.1124788

[83] Anja Thieme, Madeline Balaam, Jayne Wallace, David Coyle, and Siân Lindley. 2012. Designing wellbeing. In *Proceedings of the Conference on Designing Interactive Systems 2012 (DIS '12)*. ACM, New York, NY, USA, 789. DOI:http://dx.doi.org/10.1145/2317956.2318075

[84] Ryan Tonkens. 2009. A Challenge for Machine Ethics. *Minds and Machines* 19, 3 (July 2009), 421–438. DOI: http://dx.doi.org/10.1007/s11023-009-9159-1

[85] James Vincent. 2017. Transgender YouTubers had Their Videos Grabbed to Train Facial Recognition Software. https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset. (2017). Accessed: 2018-12-17.

[86] Yilun Wang and Michal Kosinski. 2017. Deep Neural Networks are More Accurate than Humans at Detecting Sexual Orientation from Facial Images. *Journal of Personality and Social Psychology* 114, 2 (2017), 246–257. DOI:http://dx.doi.org/10.1037/pspa0000098

[87] Jenny Waycott, Amee Morgans, Sonja Pedell, Elizabeth Ozanne, Frank Vetere, Lars Kulik, and Hilary Davis. 2015. Ethics in Evaluating a Sociotechnical Intervention With Socially Isolated Older Adults. *Qualitative health research* 25, 11 (2015), 1518–1528. DOI: http://dx.doi.org/10.1177/1049732315570136

[88] Allison Woodruff, Sarah E Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. 2018. A Qualitative Exploration of Perceptions of Algorithmic Fairness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2018 (CHI '18)*. ACM, New York, NY, USA, 1–14. DOI: http://dx.doi.org/10.1145/3173574.3174230

[89] Niels Wouters and Frank Vetere. 2019. Holding a Black Mirror up to Artificial Intelligence. https://pursuit.unimelb.edu.au/articles/holding-a-black-mirror-up-to-artificial-intelligence. (2019). Accessed: 2019-03-21.