The Work to Make Facial Recognition Work

CHRISTIAN GREIFFENHAGEN, The Hong Kong Polytechnic University
XINZHI XU, The Chinese University of Hong Kong
STUART REEVES, University of Nottingham

Facial recognition technology (FRT) has become a significant topic in CSCW owing to widespread adoption and related criticisms: the use of FRT is often considered an assault on privacy or a kind of neo-phrenology. This discussion has revolved around uses of FRT for identification, which are often non-voluntary, in particular for surveillance wherein people are (by and large) unwittingly recognized by FRT systems. At the same time, we have also seen a rise of forms of FRT for verification (e.g., passport control or Apple's Face ID), which typically are overt and interactive. In this paper we study an interactive FRT system used for guest check-in at a hotel in China. We show how guests and bystanders engage in 'self-disciplining work' by controlling their facial (and bodily) comportment both to get recognized and at times to avoid recognition. From our analysis we discuss the role of preparatory and remedial work, as well as dehumanization, and the importance of CSCW paying closer attention to the significance of interactional compliance for people using and bystanding facial recognition technologies.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI);** *Empirical studies in HCI*

KEYWORDS

face recognition; facial recognition technologies; biometric; identification; verification; ethnomethodology

ACM Reference format:

Christian Greiffenhagen, Xinzhi Xu, and Stuart Reeves. The work to make facial recognition work. *Proc. ACM Hum.-Comput. Interact.*, CSCW, Article TBD (Nov 2022), 29 pages. https://

1 INTRODUCTION

Facial recognition technology (FRT) has generated huge debate within a broad swathe of research communities (e.g., CSCW, HCI, AI, Social Sciences, Law). Much of this discussion has understandably focused on concerns about what some have labelled as non-voluntary [1] uses of facial recognition, particularly for surveillance [11, 12, 23, 56], but also other kinds of analysis, e.g., of emotion [61], gender [51], race [24], or disability [6]. The signature feature of such non-voluntary systems is that they purposefully *exclude interactivity* with persons being recognized. Our focus is on a slightly different, but nevertheless complementary phenomenon in CSCW's research interests: interactive facial recognition, which tends to be used as a method of *verification* [23] (e.g., 'face unlock' features on mobile phones, passport control, etc.). The design of such FRT

Author's addresses: Christian Greiffenhagen, Department of Applied Social Sciences, The Hong Kong Polytechnic University, christian.greiffenhagen@polyu.edu.hk; Xinzhi Xu, Department of Sociology, The Chinese University of Hong Kong, xinzhi.hsu@link.cuhk.edu.hk; Stuart Reeves, Mixed Reality Lab, The University of Nottingham, stuart.reeves@nottingham.ac.uk

XXX/XXX/XXX
https://doi.org/XXX

systems adopt a more clearly *interactive configuration* between the facial recognition system and their users (although it's important to recognize that such interactions can nevertheless be optional or mandatory [1] and thus potentially coercive, e.g., exam proctoring software [60]). In this paper, following the approach of [8] and [27], we study the 'work' that is involved in making facial recognition 'work', i.e., making oneself recognizable by a machine, both for direct 'users' and proximate bystanders who are party to FRT in use.

While interactive facial recognition has been a topic within HCI (e.g., [15]), often with a focus on usability gains (e.g., Apple Face ID), the *interactional, social organizational* consequences of facial recognition—that is, what happens in the materiality of encounters with interactive FRT—has seen little attention both in CSCW—where we might expect an stronger interest in the role of social organization around FRTs—and HCI. Our observations in this paper shine a light on the 'self-disciplining work' needed to get interactive facial recognition to work: that is, the production of appropriate forms of bodily and facial compliance. To underpin this we draw inspiration from Foucault's conceptual work on 'self-discipline' and 'technologies of self'', which modern individuals have adopted to transform their bodies so as to conform to certain ideas and norms. Foucault describes how such technologies treat the body as a machine in that the aim is often to optimize the body in various ways [18, 19]. We also trace a connection between our use of 'disciplining' and Tekin's empirical work on Kinect games which explores disciplined bodies in skillful play [58, 59],

A lot of such self-disciplining work is done through remedial work. For non-voluntary systems, remedial action is difficult to do in-the-moment. For example, in cases of wrongful arrest based on facial recognition [31], it is difficult to launch an immediate protest owing to the ways recognition is hidden, performed behind the scenes. In contrast, interactive facial recognition creates situations where getting recognized / verified is not just *accomplished* in the moment, but can potentially be *repaired* in the moment. The reason for this is that the phenomenon of verification is fundamentally constructive: as pointed out in the case of passports, the "objective practices of verification [of identity]" actually "produce the very criteria they utilize" [50]. In other words, "verification produces the verifiable object" [50]. Facial recognition technologies—like all technologies of recognition—rely upon the social production of 'self-disciplining', which is exactly what we unpack here.

Our paper thus builds upon present interests in interactive FRT applications. In doing so, we propose interactional compliance and disciplining work as new concepts for understanding the nature of facial recognition and its role in social organization. Interactive facial recognition is considered desirable for users, as its deployment happens by way of convenience rather than control. Simultaneously, it may be that the generation of facial recognition data alone—no matter the proposed usability gains—may nevertheless be problematic, contributing to models that then can be applied elsewhere, and that the argued benefits (e.g., easier usability, etc.) do not outweigh the downsides [56]. Equally, interactive FRT is not somehow immune to the well-known harms of facial recognition. Nevertheless, our aim is to surface some of the material practices that constitute interactive facial recognition in-the-world.

To explore interactive facial recognition, we captured video recordings of an automated hotel check-in kiosk that uses facial recognition to verify guests against an existing identity database that contains face data. Our data was obtained at a hotel chain situated in China which has adopted a live facial recognition system as a way of supplementing existing guest services and providing self-service capabilities. Examining these recordings in detail enabled us to understand how and in what ways 'self-disciplining' emerges as a key feature of guest encounters with FRT, in this instance for authentication purposes. However, the data may provide useful lessons about facial recognition in other circumstances.

In the next section, we briefly review and summarize some key distinctions for facial recognition that situate our study, examine notions of 'the face' at play in them, and explore common critiques of FRT. After presenting our setting and describing our data collection and analysis (Section 3), we introduce the basic features of how guests got recognized through facial

recognition (Section 4). We then show how guests *prepare* themselves for the encounter (Section 5), and may engage in *remedial* work if there are problems with the recognition (Section 6). In the discussion (Section 7), we reflect on the significance of compliance and disciplining work in facial recognition, the potential for dehumanization, and discuss the implications of our study for how facial recognition might be further investigated in CSCW research.

2 FACIAL RECOGNITION: ANALYSIS, IDENTIFICATION, AND VERTIICATION

We want to begin by differentiating different ways in which FRT can be used. For our purposes we distinguish between three uses: for analysis, identification, and verification. To summarise, analysis is about using the face as an index for 'the inner', identification is about individuating people, while verification seeks to match a representation of an individual with a given face.

2.1 Analysis: the face as a 'window to the soul'

The face is often seen as endowed with meanings of different sorts; for example, eyes are often seen as 'windows to the soul'. In that sense, the face is viewed as beyond our control (in contrast to, for example, the words we use) and therefore analyzable as representative of one's 'inner states'. This conception has led to attempts at various kinds of analyses of the face that seeks to extract information deemed available from facial input. This information includes the presumed attributes of faces like gender [34, 51], the messages that faces convey, such as emotion or mood [5], recognized via expression analysis [23]. The first use of FRT is thus facial *analysis*, which seeks to read off such information from the face, working on the assumption that phenomena as complex and disparate as one's gender, emotions, or even truthfulness are primarily the domain of the face.

There is, however, another sense to the face in that it is also seen as representing an 'index to identity' [23, p. 8]. Individualization in modern (Western) societies has rendered the concept of the individual as 'natural', however the idea is a modern one that is technologically mediated. Accordingly, it has become crucial to specify and verify someone's identity[1] in different situations and for various purposes. Technically this identification links an input (image, video, etc.) of a specific face to a specific individual. This interest has led to 'technologies of verification' such as passports [50] and increasingly sophisticated methods of *biometric verification* (fingerprinting, iris scanning, voice recognition, DNA analysis), the aim of which is to "bind identity to the body using digital representations of unique body parts, or, in the case of voice printing, by capturing, digitizing, and analyzing the sounds that the body produces" [23, p. 14]. To the development of such biometrics, we now must add new facial recognition technologies, where "the face becomes the new fingerprint" [26, p. 127] and thus a key resource for identification.

When FRT is using the face as an index for identity we can broadly distinguish (following [23, p. 18] between a use for *identification*, where FRT is used to identify persons whose identities are presently unknown (e.g., locating someone in a crowd) and a use for *verification*, which attempts to check whether individuals are who they claim to be [23].

2.2 Identification: non-voluntary facial recognition

When FRT is used for identification, this is often non-voluntary. This includes applications of CCTV to discretely identify specific people against a database of faces [23, Chapter 2], perhaps to track suspected criminals or people sought by immigration officials [1, p. 140]. Such non-voluntary uses effectively represent a *surveillance* application of FRT, leveraging networks of input data to expand coverage.

As we saw earlier, the individualization of society meant that identification has become a growing concern. In the nineteenth century, Jeremy Bentham argued that in order to prevent

[1] From our perspective identifying a person is still 'analysis' in the sense that recognising someone represents an analysis just as recognising that they are upset also represents an analysis. So our distinction is primarily for practical purposes in this paper, rather than suggesting a schema we want to sign up to.

crime it is important to identify individuals, the reasoning being that if you can disguise yourself, you feel safe enough to commit crimes (which is why people wear masks when committing crimes). Indeed, Bentham already anticipates facial recognition technologies of sorts:

> "There is a common custom among English sailors, of printing their family and Christian names upon their wrists, in well-formed and indelible characters; they do it that their bodies may be known in case of shipwreck. If it were possible that this practice should become universal, it would be a new spring for morality, a new source of power for the laws, an almost infallible precaution against a multitude of offences, especially against every kind of fraud in which confidence I requisite for success. Who are you, with whom I have to deal? The answer to this important question would no longer be liable to evasion." [7, p. 557]

The majority of research has tended to focus on such non-voluntary applications of facial recognition technologies for identification, centering on surveillance for the purposes of various societal control measures like crime detection and prevention [23], doing their part for the construction of a 'surveillance society' [39].

Not surprisingly, non-voluntary applications of FRT for identification have received substantial criticism. These range from pointing out that facial recognition for identification simply does not work (e.g., recognizing criminal suspects [31]) to highlighting various forms of biases that are incorporated into them, in particular, with respect to race and gender [2, 12, 24, 56]. Indeed, critics have pointed out that facial recognition technologies make matters of prejudice worse, because they may *appear* more neutral than humans. For example, in the context of law enforcement, they "generate many of the same biases as human law-enforcement officers, but with the false patina of technical neutrality" [12, p. 565]. Some scholars suggest bans on facial recognition technologies for almost any kind of non-voluntary application (e.g., [56]); others critique some attempts to debias AI technologies like facial recognition and so produce ethical AI, arguing that many approaches to this problem are designed to co-opt and neuter genuine challenges to the use of facial recognition [33]. As a consequence, tools to expose algorithmic bias in facial recognition systems have been proposed [45], as well as methods of resistance to recognition technologies, such as 'Camera adversaria' [10], which attempts to disrupt the automatic surveillance of personal photographs, or artist Adam Harvey's computer vision dazzle makeup, which seeks to disrupt non-voluntary uses of facial recognition technologies in public spaces [28].

2.3 Verification: interactive facial recognition

Rather than using FRT for identification, FRT can be used to verify the identity of people. These implementations do not conceal that facial recognition systems are being applied to people's faces and indeed may require more active participation by the observed individuals. Face unlock for mobile phones or hotel check-in systems [38, p. 170] are both examples of this use, demonstrating that "the face continues to operate as a territory of individualization" [11, p. 83]. Within voluntary systems, facial recognition may be mandatory to obtain some benefit or optional, where alternatives are offered [1, p. 139]. Although voluntary, one can argue such systems still contribute to potentially building large databases of sensitive biometric data, and can nevertheless still be coercive, particularly if *no* alternatives are provided (e.g., if Apple ID was the *only* way to unlock a phone); nevertheless, they certainly do not *hide* from a user that facial recognition is being performed.

While there is a large body of literature focusing on non-voluntary applications of facial recognition technologies and their problems, we find much less research investigating interactive uses of verification. Researchers point to varied levels of acceptance of facial recognition technologies and that "surveillance and control are not foremost on the minds of citizens in China, Germany, the United Kingdom, and the United States, but rather notions of convenience and improved security" [35, p. 671). Yet, similar to criticisms of surveillance-oriented uses of FRT, there are basic reliability problems when compared with the ability of other biometrics like DNA or fingerprints which themselves are not without a wide range of problems [21, p. 139; 23, p. 17].

Further, applications that employ voluntary uses of facial recognition for authentication may introduce insecurities, such as children being able to unlock a parents' phone [21, p. 140].

Moreover, studies of interactive facial recognition technologies do not discuss the work done by the people getting recognized and the ways in which they must comport themselves to pursue successful verification. This study thus complements broader critical perspectives on facial recognition technologies use that have yet to address the material, thoroughly *practical* work involved in this recognition.

3 THE SETTING

Our setting is a fully self-service chain of hotels in Guangzhou and Shenzhen, China, that had started using FRT as part of their business. The first chain of hotels was opened in 2016, being the first 'robotic' self-service hotel in China. The hotels make use of various service robots, along with automated self-registration and check-in kiosks [38]. Human service workers are generally absent from the scene unless guests encounter problems they cannot resolve on their own (the chain name translates literally as "no human service worker"). After guests have successfully checked-in using the FRT-enabled kiosks in the lobby, a separate service robot (see Figure 1, left) leads them to their rooms and unlocks the door.

The facial recognition system we examine was installed into the lobby area of the three hotels in the form of a check-in kiosk (see Fig. 1).

Fig. 1. Hotel lobby and kiosks with the facial recognition check-in system (ID card placement circled in red) and route-leading service robot (left)

When the guest approaches the kiosk, there are two buttons to choose from: "registered-person check-in" and "roommate check-in". Once they have selected one of these options, the kiosk asks them: "please place your ID card" (the message is delivered both on the screen and via loudspeakers; ID card placement indicated in Fig. 1). The process starts when the check-in machine acknowledges the successful reading of the card with an audible beep. Subsequently, the kiosk asks the guest via loudspeaker: "please remove your ID card". At this point, the screen

switches to the facial recognition interface (see Fig. 2. and Fig. 3), after which the system announces (via loudspeaker and on the screen) that it is "facial recognizing". The guest then follows a series of instructions to check into the hotel successfully. These instructions are provided in sequence as follows:

(1) "please close your eyes" (请闭眼);
(2) "open your eyes" (睁眼);
(3) "please shake your head" (请摆头);
(4) "stay still" (摆正);
(5) "please open your mouth" (请张嘴);
(6) "close your mouth" (合嘴).

At any point during this instruction sequence facial recognition may be successful, and the guest will move onto the next stage of the check-in process. The subject of our analysis lies is in the details of how guests get recognized.

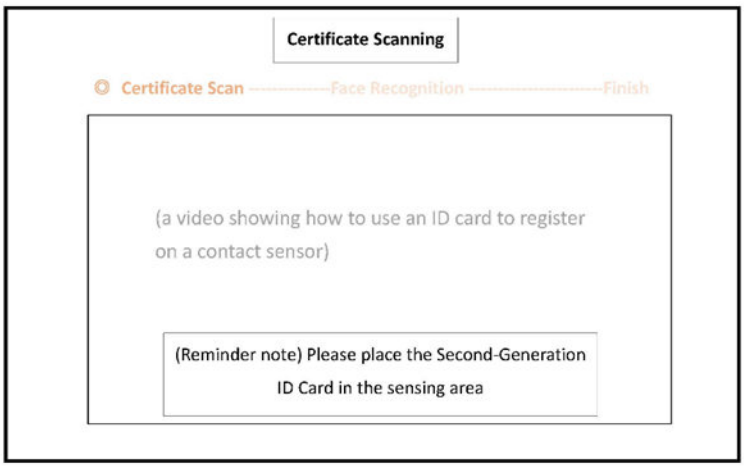

Fig. 2. The interface of 'ID card verification' of the registration kiosk

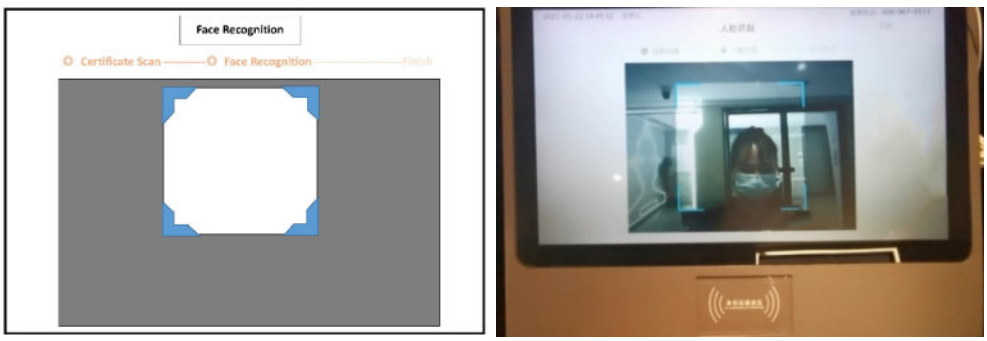

Fig. 3. Diagram of the facial recognition screen and its various parts (left) with the square in which a guest's face should be visible indicated with blue corners; photograph of what the guest sees (right).

There is an element of novelty factor to the hotels, with many guests attending because of their aforementioned 'robotic' element. Other guests may of course be visiting the hotel chain owing to its location or price. In this sense, the hotels position themselves explicitly with

automation and machinic recognition as a feature. This non-coercive aspect of FRT use here should be noted so as to properly contextualize and situate our study alongside other instances of FRT that we have previously discussed.

As hotels located in China, it is worth briefly situating them with respect to existing uses of facial recognition, as we think it likely many guests will have had prior experiences or some familiarity with facial recognition. FRT growth in China is high both in terms of new companies (e.g., more than 900 facial recognition patents were filed in 2017 as compared with 96 for the USA [51]) and projected global share (~44% in 2023) [63]. Underpinning this is a high (and growing) penetration of CCTV cameras into everyday life [3]. Facial recognition capabilities are routine features for many of the new cameras being produced [51]. This is atop many existing forms of biometric identification in use in China, which FRT acts as an extension of. For instance, increasing numbers of commercial organizations use FRT, it is often applied in accommodation access control, and routinely for security in banks, to name a few [63]. With respect to this situation, it has been argued that biometric technologies like FRT are convenient and offer improved security [35, p. 671; 63], but we also find mixed research on acceptance by Chinese publics—some suggest FRT is widely accepted [35, p. 671] whereas others report that it is in reality strongly rejected by residents [8]. There are also signs of movement in terms of government regulation 64, and increasing public contestations of FRT use [66].

4 STUDY DESIGN

We conducted fifteen months of fieldwork (between 2019 and 2021) in the three aforementioned hotels. During this time we collected video recordings of interactions with / around the check-in kiosks, resulting in 78 hours of video, with several hundreds of guests. Video capture was spread between the three hotels in the following way: six months was spent in Hotel A, Shenzhen, five months in Hotel B, Shenzhen, and four months in Hotel C, Guangzhou. The size of each hotel is around 50 rooms.

To conduct this fieldwork, one of the authors (Xinzhi Xu) actually worked as a part-time intern in all three hotels to familiarize herself with the setting, learn the routines of interaction, and as part of this also gain access for conducting video recording of the kiosks in use. And at the beginning of the data collection period, the hotel managers approved an initial pilot recording conducted during which the researcher was always present. This was to assess whether guests felt uncomfortable or found the consent approach problematic. After this was deemed appropriate, the researcher began recording for longer periods of time during which she was sometimes away from the site, or on-site working as a part-time intern at the hotels as we previously mentioned. The length of the fieldwork was determined by the uncertainty about when guests would check-in at the hotel and the need to collect as many cases of interaction given this.

We gained permission for capturing video recordings in all three hotels. Following previous studies that used video-recordings in public settings, we adopted an 'opt-out' model of consent for customers [30, p. 25]. We posted prominent signs next to and on the kiosks themselves (in both Chinese and English), which described the purpose of the project, contact details for the researcher, and information about the video-recording, including that we might use non-anonymized images in publications [30, p. 30]. The signs contained clear instructions for opting-out of the study, locating the researcher on the spot, verbally indicating out loud to the camera by saying "I don't want to participate in this project", or performing a physical gesture to indicate non-consent in front of the camera to opt-out. All cases of minors being captured (3 in total) were excluded from our collection. Note that the researcher was often present during video capture periods to facilitate consent. The 'opt out' method for both the video-recording and the use of non-anonymized images were approved by <university research ethics committee>. Much of our data is to do with small facial changes and movements, and thus anonymization such as blurring faces would render many of our examples difficult to understand and hinder the value of our study.

We took an ethnomethodological (EM) perspective to the phenomenon of interactive FRT use, which means unpacking how encounters with interactive facial recognition are socially organized. EM—and relatedly, conversation analysis (EMCA)—has a long history in CSCW of developing our understanding of how digital technologies come to be embedded in and as practices of routine social organization [57]. Our study takes place in a customer-based and service context; various related studies have looked at how technologies get used in these circumstances that adopt an EM approach. For instance, [29] looks at collaboration work between, and amongst, managers and the control system in public transport control rooms, [65] details how technicians use index menus to locate customers' problems in software, while [42, 43] explores how clerks make customer's unique requirements accessible for workers who operate on a photocopy machine. A further relevant interest for ethnomethodological studies which our EM approach draws upon is in examining the social organisation of body movements, including facial actions. For instance, [14] and [16] note how the body and face become resources in video calls, [36] identifies how 'talking head' arrangements in video calling act as a 'default mode' (to which actions are accountable to), while [22] shows that a child's gaze is a preconditional matter for connections during children-parent video calls.

EM is often understood by some as a 'qualitative' methodology[2], where researchers typically present a small number (four or five) of perspicuous examples to highlight the key characteristics of the phenomenon under investigation. Our approach in this paper is slightly unusual in this regard, since we also present some more 'quantitative', numerical information. Since the process under consideration is highly recurrent, we ended up with a large number of cases (in particular, 361 of successful facial recognition, see Section 5.3 below). We therefore sometimes present some basic descriptive statistics (e.g., a distribution of the number of cycles of instructions it took guests to get recognized, see Fig. 1Fig. 6). However, we want to emphasize to the reader that this should be understood ultimately as *scenic*, background context to the core 'qualitative' analysis of key instances.

5 GETTING RECOGNIZED

To begin, we detail how the facial recognition process starts for guest check-ins by examining an instance where the guest is quickly successful—that is they get recognized rapidly (within seconds) and encounter no visible trouble. Following this example, we then explore a case where recognition takes a long time, which help us to understand how things may be troublesome for people interacting with FRT. To further contextualize these examples, we will then provide an overview of ways guests succeeded or failed to be recognized.

5.1 Quick success

Here a guest (GUE) checks in successfully within a few seconds of following the cycle of instructions from the check-in kiosk (KIO); see Fig. 4. We join the fragment just as the guest has placed and withdrawn his ID card from the sensor (the kiosk displays an interface as Fig. 3); the facial recognition system starts at line 02 (we have included a video file 'quick-and-slow-success' for reviewers, which can be accessed through the submission website). In Fig. 4 we use a standard form of transcription common in conversation analysis [31].

[2] As ethnomethodologists we would subscribe neither to the distinction between qualitative and quantitative, nor consider ethnomethodology a 'method', however for the purposes of this paper we gloss EM as 'qualitative'.

((the guest has placed and withdrawn his ID card from the sensor;
facial recognition starts at line 02))

01	GUE:	((looks at the screen))
02	KIO:	人 脸 识 别 中 *ren lian shi bie zhong* *facial recognizing*
03	KIO:	请 闭 眼 *qing bi yan* *please close your eyes*
04		(0.4)
05	GUE:	(0.4) // ((closes eyes hard))
06	KIO:	[睁 眼 *[zheng yan* *[open your eyes*
07	GUE:	[((keeps eyes close))
08	GUE:	(1.2) // ((opens eyes))
09	KIO:	请 摆 头 *qing bai tou* *please shaking your head*
10	GUE:	((shaking head))
11		(2.0) // ((the screen turns to the next interface))



Fig. 4. An example of a 'quick success' in achieving facial recognition

Some basic features of the system are clear in this example. The kiosk indicates audibly to the guest that face recognition is taking place (line 02), which is swiftly followed by the first instruction: "please close your eyes" (line 03). After a short pause the guest responds by closing his eyes (line 05), keeping them closed until the instruction (line 08) to open them. Then the next instruction arrives, asking him to shake his head, which he does on line 10 (this time with little to no pause). The kiosk, after a two second pause, changes the screen to the next stage of the check-in process (entering a phone number and sending a subsequent verification code to the phone). The whole sequence takes nine seconds. In other words, the guest succeeds in the facial recognition process after only three instructions ("close eyes", "open eyes", "shake head") and, more importantly, before the final instructions ("stay still", "open mouth", "close mouth"). In line 11, the screen turns to the next interface, which requires the guest to input their mobile phone number to get a verification code.

5.2 Failure

This interaction is not the whole picture, of course. Interactive facial recognition fails for guests from time to time in a variety of ways. In the following example (Fig. 5), another guest (GUE) is trying to check-in at a kiosk (KIO) and ends up going through the three cycles of instruction pairs (i.e., being asked to open/close eyes, shake head/keep still, open/close mouth) until finally (after 38 seconds) the machine reports: "facial recognition failed, please try it again". The guest then must start the process again. The guest then succeeds in being recognized after the instruction "please shake your head" during the first cycle of the second attempt (A2C1) (this is the second part of the video file 'quick-and-slow-success', which can be accessed through the submission website):

Attempt 1, Cycle 1 (A1C1)	**Attempt 1, Cycle 2 (A1C2)**	**Attempt 1, Cycle 3 (A1C3)**	**Attempt 2, Cycle 1 (A2C1)**
[00s] KIO: "face recognizing"			[52s] KIO: "face recognizing"
[02s] KIO: "please close eyes"	[14s] KIO: "please close eyes"	[25s] KIO: "please close eyes"	[54s] KIO: "please close eyes"
[03s] GUE: ((blink))	[15s] GUE: ((blink))	[26s] GUE: ((close eyes))	[55s] GUE: ((close eyes))
[04s] KIO: "open eyes"	[15s] KIO: "open eyes"	[27s] KIO: "open eyes"	[56s] KIO: "open eyes"
[05s] GUE: ((blink))	[16s] GUE: ((blink))	[28s] GUE: ((open eyes))	[57s] GUE: ((open eyes))
[06s] KIO: "please shake head"	[17s] KIO: "please shake head"	[29s] KIO: "please shake head"	[58s] KIO: "please shake head"
[07s] GUE: ((shake head))	[18s] GUE: ((shake head))	[30s] GUE: ((shake head))	**[63s] ((facial recognition success))**
[08s] KIO: "stay still"	[19s] KIO: "stay still"	[31s] KIO: "stay still"	
[09s] GUE: ((shake head))	[20s] GUE: ((stay still))	[32s] GUE: ((stay still))	
[09s] KIO: "please open mouth"	[20s] KIO: "please open mouth"	[32s] KIO: "please open mouth"	
[10s] GUE: ((open mouth))	[21s] GUE: ((open mouth))	[33s] GUE: ((do nothing))	
[11s] KIO: "close mouth"	[22s] KIO: "close mouth"	[34s] KIO: "close mouth"	
[12s] GUE: ((close mouth))	[23s] GUE: ((close mouth))	[35s] GUE: ((do nothing))	
		[38s] KIO: "facial recognition failed, please try again"	
		[39s] KIO: ((change interface))	
		[40s] GUE: ((swearing))	
		[43s] GUE: ((operate on KIO))	
		[45s] KIO: "please place ID card"	
		[46s] GUE: ((place ID card))	
		[48s] KIO: "please remove"	
		[49s] GUE: ((remove the ID card))	

Fig. 5. An example of repeated 'failure' in achieving facial recognition (numbers on left indicate the time in seconds); GUE=guest, KIO=kiosk

In our first example (Fig. 4), the guest is successfully recognized within the first cycle, never reaching the "please open your mouth" instruction, and not even encountering the second part of the "shake head" / "stay still" pair. In contrast, in the second case the guest spends 40 seconds trying to get recognized by the check-in machine, frequently following the instructions. Ultimately the attempt fails and the machine announces through a loudspeaker: "facial recognition failed, please try again". The guest then starts a second attempt (fourth column, A2C1) and eventually achieves successful facial recognition 63 seconds after starting.

5.3 Summary: Success and failure at getting recognized

In order to provide readers with a sense of how quickly guests get recognized in this setting, we can provide some basic statistical information as a background for our qualitative analysis.

Our data contains 412 cases of guests attempting to use the automated check-in process. In 28 cases, the facial recognition system never starts due to problems with ID cards (e.g., the kiosk cannot read the card, leading to manual check-in with a service worker). In 23 further cases, there is a system error of some sort e.g., "no available room" or "can't find reservation", and although the facial recognition system starts, guests do not encounter any instructions.

We are left with 361 cases in which guests achieve successful recognition[3] (see Fig. 6). Only in 79 cases of these 361 (22%) do the guests get recognized during the first cycle of instructions. In all other cases, the guests encounter the instructions multiple times. For these 282 cases (78%) we can broadly distinguish between two situations. On the one hand, a situation where guests achieve success *before* the machine produces an explicit failure notification, i.e., "facial recognition failed, please try again" (see Fig. 6), which happened in 230 cases (64%) where guests get recognized during the second (A1C2) or third cycle (A1C3). On the other hand, a situation where the success is achieved after the failure notification, which occurred in 48 cases (13%) where guests get recognized on the second, third or fourth attempt, after encountering one or more failure notification.

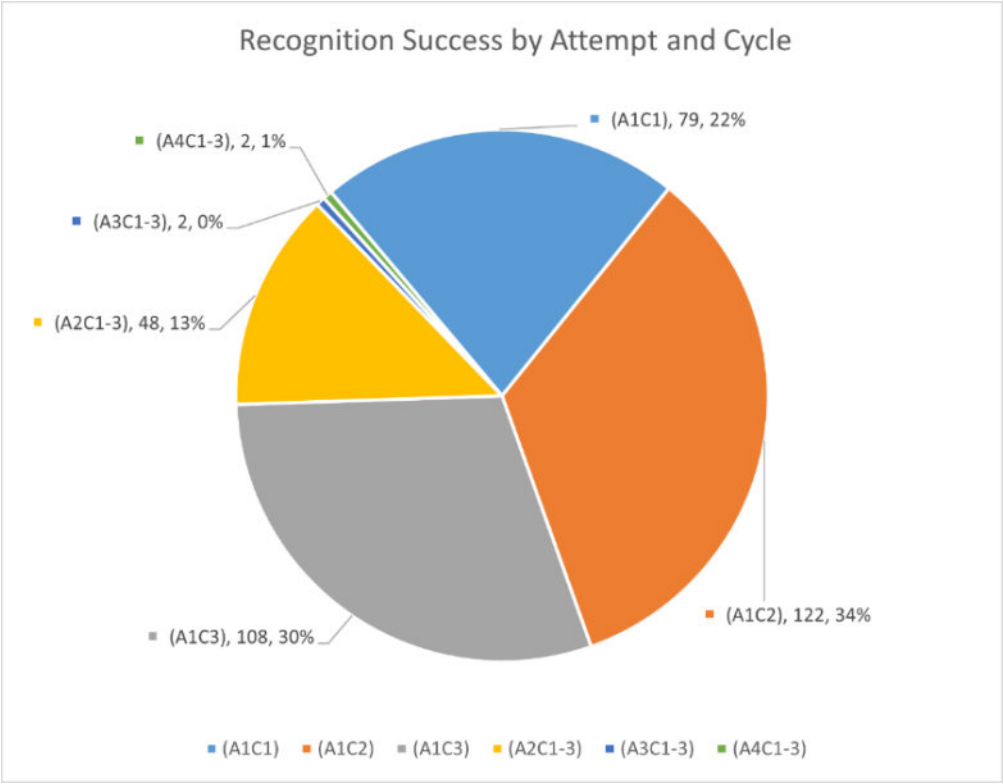

Fig. 6. When guests succeed with recognition, broken down by attempt and cycle

Quick success (as shown in our first example) is rare, but so is very delayed success (as shown in our second example). Most of the action thus resides somewhere 'in between'. These results suggest a mixed picture of interactive facial recognition, which can involve a non-trivial amount of practical work on the part of the guests to get the facial recognition working and thus get recognized. It is precisely this work that we are interested in exploring for the remainder of our paper. Indeed, in our data, this work often quite noticeably starts before any FRT has been initiated.

The nature of this work is also particular to interactive facial recognition. Unlike most of the facial recognition literature in HCI, CSCW and beyond, which tends to study non-voluntary facial recognition where it is perhaps more natural to actively *avoid* FRT (if they are, say, tied up with surveillance), in our case people actually want to work with facial recognition to get verified and thus checked-in. At first this will seem to tell us little about the ethics of facial recognition as

[3] In about 70 of these 361 cases, the service worker gets involved by helping the guest (see Fig. 13.).

other studies often have elaborated at length. However, it will lay out our initial emphasis on the practical work necessary to make facial recognition 'do' its apparent feat of recognition—practices that tend to be left out of accounts of facial recognition deployments 'in the wild' but that are nevertheless are likely prevalent.

6 PREPARATION WORK

Before following the instructions of the machine, guests already *pre-empt* what might be necessary for getting recognized. We might say that guests are prospectively *making themselves machinically recognizable*. Here we outline a few of the ways that this anticipatory orientation towards the needs of the facial recognition system manifest in and as the actions of guests—including those acting as bystanders. These processes include guests attending to spatial position, the wearing of particular accessories, and bystander sensitivity towards both the unfolding interaction between the checking-in guest and the kiosk, as well as their own potential for (undesired) interactions with it.

6.1 Adjusting position

Guests often adjusted their physical position in the hotel foyer to manage their availability for facial recognition. Many of these adjustments seemed to be oriented by feedback from guests' visual appearance in the transparent viewing frame presented on the screen (see Fig. 3, left). Stepping this way or that enables an optimum alignment for recognition to take place. Fig. 7 is one such example.

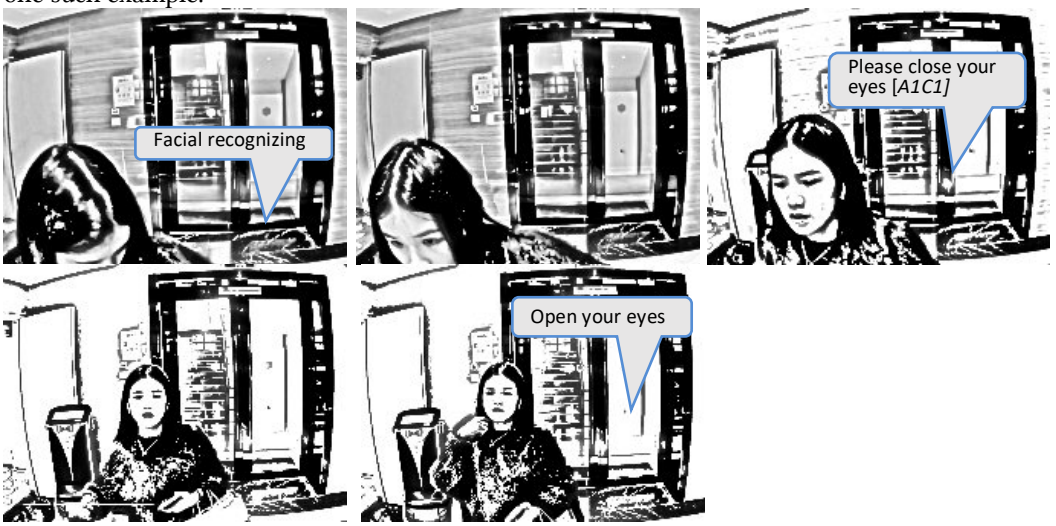

Fig. 7. Guest takes three steps backwards after the initial announcement from the kiosk

After the initial announcement from the kiosk that it is "facial recognizing", this guest takes a number of steps backwards. The guest's movements then continue after the first instruction ("please close your eyes") as they continue their trajectory towards what seems to be their desired spatial location for their interaction with the facial recognition system. We could say that the guest is staging the position of their face such that it appears 'formatted' readily, compliantly, for the facial recognition technologies to process.

6.2 Removing accessories

Guests' concern for arranging their bodily position in a particular space could in part be explained by their visual on-screen appearance. Yet we found anticipatory compliance went well

beyond this feedback required. Displaying a sensitivity towards, and simultaneously also an analysis of, facial recognition technologies, guests often pre-emptively removed worn accessories, seemingly oriented by their own analysis of what constitutes a recognizable face. This anticipatory removal of accessories to avoid potential trouble in getting recognized suggests guests are oriented towards certain mundane reasoning about machine recognition (which may be related to levels of public awareness of FRT in China).

Fig. 8. shows three such examples.

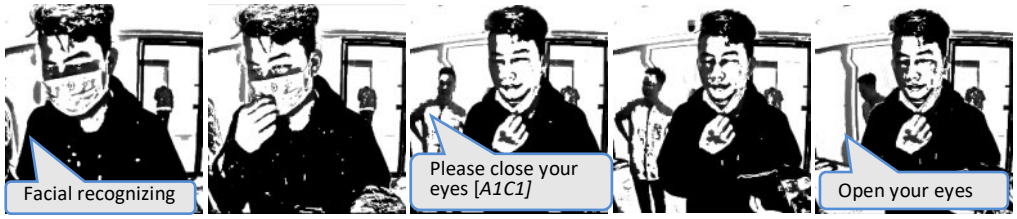

Fig. 8. Guests removing hats (top), glasses (middle), and masks (bottom) on or before the first instruction

Guests are not explicitly instructed on exactly how to prepare their face to enhance machine recognizability, yet they often performed such actions before or alongside the first instruction. This behavior is a kind of anticipatory compliance.

We also found that 'bystanders'—individuals not there by happenstance but rather there as roommates—also displayed co-orientation to ways that faces might be prepared for the assumed needs of recognition via the management of accessories. In the following instance (Fig. 9.), the bystanding roommate reminds the checking-in guest of the fact they have their glasses on at a specific point subsequent to the first instruction ("please close your eyes").

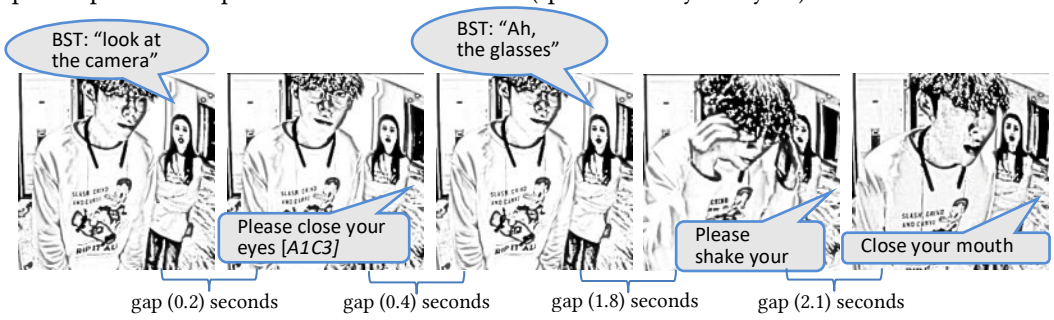

Fig. 9. Roommate reminding registering guest to look at camera and suggesting they remove glasses

This analysis by the bystander ("ah, the glasses") highlights how *others*—besides the person being recognized—can also be (co-)oriented towards potential troubles that may be encountered with facial recognition technologies, anticipating (and suggesting) remedies. Notably, the bystander was already instructing the guest prior to this interjection about the glasses. They immediately direct the checking-in guest to "look at the camera", a command subsequent to the "facial recognizing" alert from the kiosk. On the instruction "please close your eyes" the bystander then (without explicitly articulating that they may be troublesome) highlights "the glasses" as a potential stumbling block for facial recognition. The key point here, which we also explore next, is that the ostensibly 'main user' of facial recognition is not necessarily the be-all and end-all: the preparatory work done to render one's 'amenable' (or, perhaps, 'docile') to the needs of facial recognition technologies can extend to proximate others too, people who may attend to the situation continuously, providing timely directives to the checking-in guest that represent their own analysis of the situation.

6.3 Bystanders 'removing' themselves

The bystanding roommate may also engage in their own preparatory work, sensitive to producing compliance with the perceived needs of the facial recognition system. Fig. 10. shows two examples of spatial repositioning of bystanders, while the main guest continues the registration process.

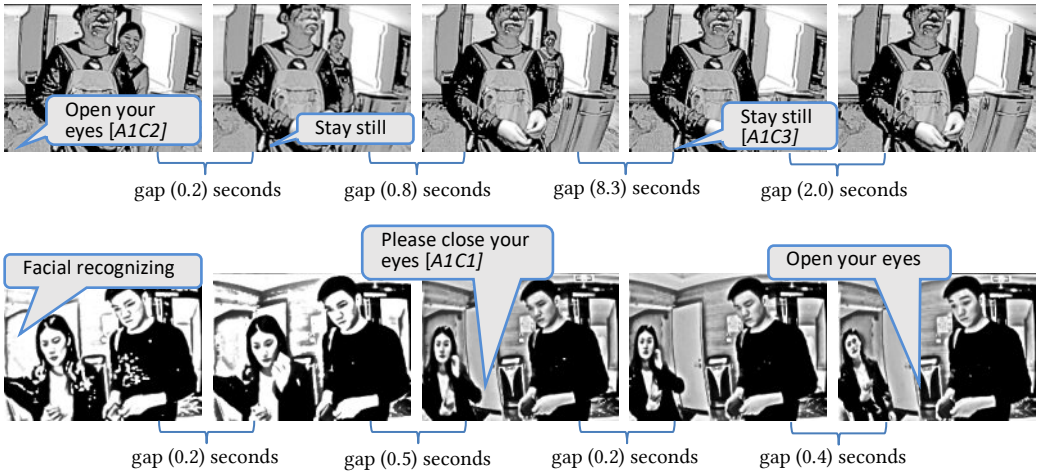

Fig. 10. Bystanding roommate repositioning themselves 'out of sight' of facial recognition kiosk cameras

Anticipating next section's focus on remedial work, we note that a bystander may be asked or directed by the guest checking-in to perform actions to handle prospective or present troubles with facial recognition. The following example is of the latter type, where the guest tells the bystander: "move move, move away".

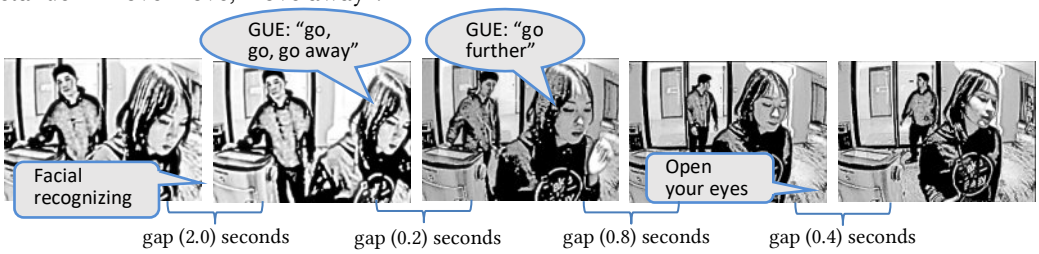

Fig. 11. Guest telling a bystanding roommate to move away to avoid being 'seen' by the kiosk's camera

In this example (Fig. 11.), after the kiosk has informed the guest that it is "facial recognizing", the guest anticipates that having someone else 'in the picture' may be a problem, which is why she instructs her bystanding roommate to "go away". Indeed, after the bystander does so, the guest checking-in seems to think that he hasn't gone away far enough and instructs him to "go further".

This final example has presented a guest's attempt to resolve a problem in response to indications of trouble emerging from their interaction with the facial recognition system. This behavior is broadly what we call 'remedial work': actions that are less about anticipating prospective problems and more about managing them as the facial recognition procedure takes place.

7 REMEDIAL WORK

In a previous section we introduced the issue of failure as reported by the machine—via explicit error reports from the kiosk that "facial recognition failed" or implied troubles where, for example, a guest hears the same instruction for a second (or third) time. As we will show, this repetition is often taken as evidence of a problem, the resolution of which is finding an answer to 'what can I do *differently*?'. We will call this work of producing different actions the 'remedial

work' conducted by guests, behavior undertaken to remedy troubles as they are encountered. This remedial work expands what we have described as 'preparation work', building on the example at the end of the last section.

7.1 After explicit failure

The clearest cases of remedial work—doing things differently—on the part of guests emerged immediately after a message of failure from the kiosk at the end of the third cycle of instruction pairs (see Fig. 5 for an overview). Just as guests displayed their own anticipations (i.e., analysis) of what it means to render themselves and their face 'machine recognizable' in and through their various productions of 'compliance'[4] above, we see similar remediates being made by guests *in response* to a failure message (including removal of accessories, adjusting spatial positioning, and so on). Fig. 12. provides three examples.

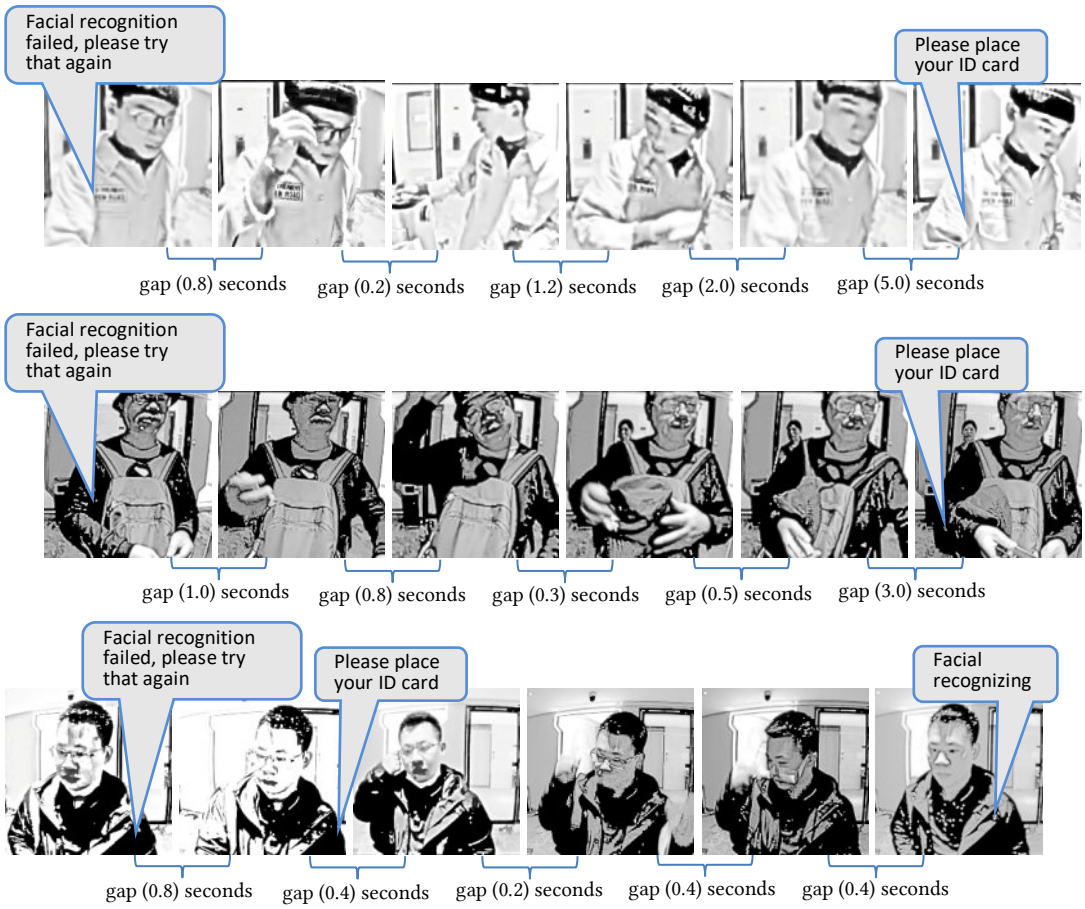

Fig. 12. Guests removing glasses (top), hat (middle) and mask (bottom) after failure message

[4] It remains unclear to what extent guests' ID card image plays a role in the ways of rendering their face more machinically recognizable. For instance, is it the case that guests align their appearance with the ID card (registered with the facial recognition database against which their face is to be compared)? If they removed their glasses for the ID card picture, does this impact the ways they attempt to resolve trouble with facial recognition? Or are guests primarily working towards compliantly producing 'recognizability' based on their appearance on-screen here-and-now appearance only? These are unresolved questions.

Indeed, such remedies are typically what proximate hotel service workers suggest after a failure, encouraging guests to remove any accessories worn on the guest's head (e.g., "take off your glasses" in three cases, see Fig. 13.) or to adopt a different physical position (e.g., "step backward a bit" in five cases, see again Fig. 13.). We found that they would also sometimes recommend a guest "try the other machine" (seven cases).

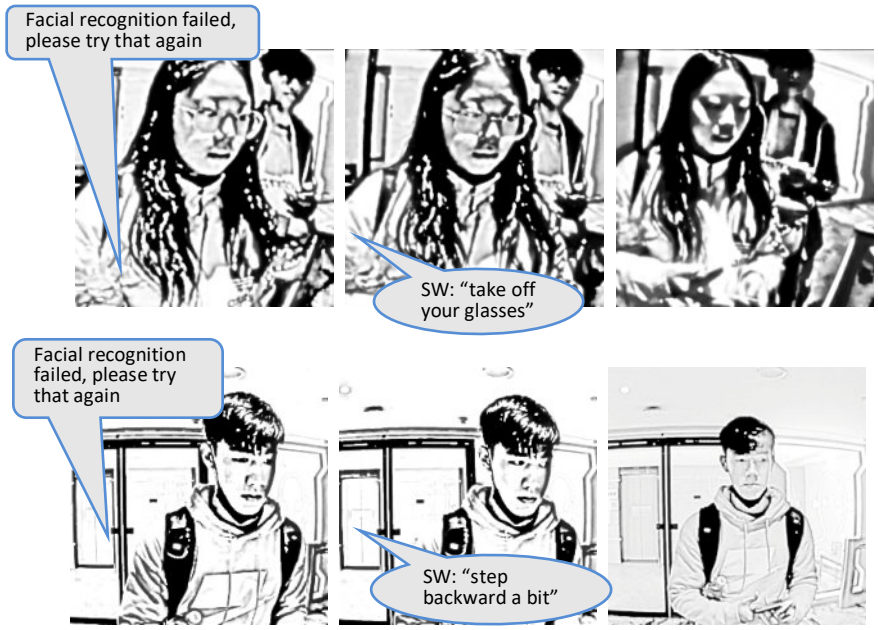

Fig. 13. Service worker (SW) suggesting ways of resolving trouble with facial recognition

As we have mentioned, most guests in our data were recognized within the first cycle, and so never encountered an explicit failure message. However, even within this sample, *repetition* could sometimes be treated as troublesome and thus occasion some remedial action by guests.

7.2 Encountering repeated instructions

Guests know that there is a problem when they are told by the machine that "facial recognition fails". However, they can become aware of the problem *before* that moment, namely when they hear an instruction for a *second time*. This awareness is something that we are all familiar with: if someone repeats the question or instruction after an initial response, it indicates that the initial response was in some way inadequate. For example, in classroom interaction, a teacher's "question-repeats are also heard as marks of a failed answer" [40, p. 719]. Thus, when a guest hears an instruction (e.g., "please close your eyes") for a second (or third) time, they can assume that there is problem. However, even though they know *that* there is a problem, they do not know *what* the problem is.

On the one hand, guests can engage in the kind of remedial work that we have already discussed: adjusting their standing position or removing accessories (see Fig. 14.).

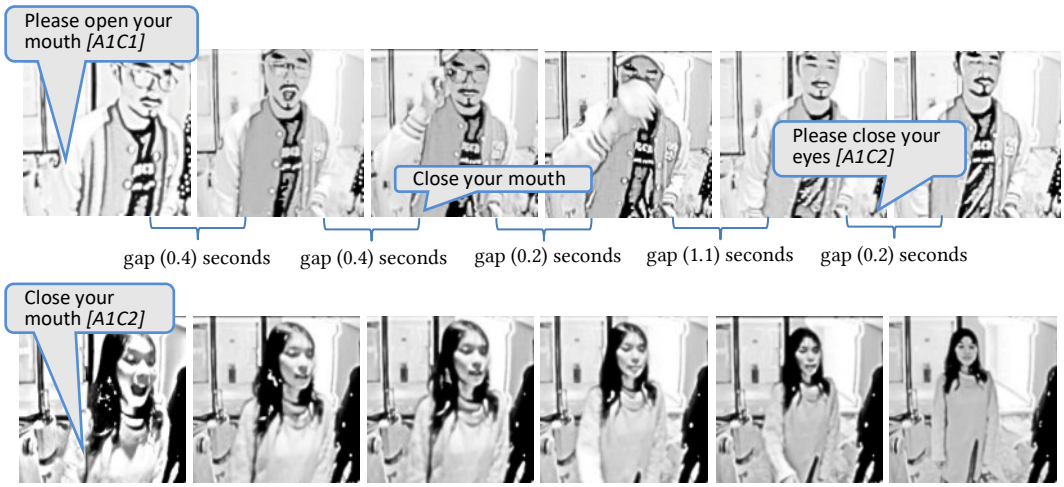

Fig. 14. Guests removing glasses and repositioning themselves towards end of the first cycle

On the other hand, guests can change *how* they respond to the instruction. Again, this response is something very familiar to all of us. When teachers repeat their question, students not only know that there is a problem with their answer, but they must *modify* their answer. Similarly, in the context of parent-child interaction, [17] observes: "By repeating their questions the parents are making explicit that another action is required."

7.3 Creating different responses

When guests become aware that there might be a problem with their response to the instruction, they may change it. Here we'll focus on how this modification manifests itself in guests' response to being instructed to close their eyes and to shake their head.

When guests are instructed to close their eyes, they can change *how long* they close their eyes and, relatedly, *when to* stop closing their eyes, that is, when to open their eyes. While guests may initially only close their eyes very briefly, in a subsequent response they may close their eyes longer and, furthermore, keep them shut until they hear the instruction "open eyes" (in the next section, we will discuss the 'paired' nature of these two instructions). Thus, in the following example shown in Fig. 15, the guest initially closes his eyes for only 0.3 seconds, and opens them *before* the kiosk tells him: "open your eyes". In the second cycle, the guest closes his eyes much *longer* (1.1 seconds) and waits to open them until he hears the instruction "open your eyes".

Kiosk	Guest actions (A1C1)		Guest actions (A1C2)	
"Please close your eyes"	((open))		((open))	
(0.1)	((open))		((open))	
(0.1)	((open))		**((close))**	
(0.1)	((open))		**((close))**	
(0.1)	((open))		**((close))**	
(0.1)	**((close))**		**((close))**	
(0.1)	**((close))**		**((close))**	
(0.1)	**((close))**		**((close))**	
(0.1)	((open))		**((close))**	
"open your eyes"	((open))		**((close))**	
(0.1)	((open))		**((close))**	
(0.1)	((open))		**((close))**	
(0.1)	((open))		((open))	

Fig. 15. Instructions generated by the kiosk alongside concurrent actions by the guest checking-in, comparing two separate cycles: first cycle (C1, left) and second cycle (C2, right)

We can think about the issue presented to guests when hearing repeat instructions in the following way. When you hear "close your eyes" for a second time, but you closed your eyes the first time, you are faced with a puzzle: What could you have done wrong?

Apart from changing the length and timing of the closing of their eyes, some guests changed the *intensity* of closing their eyes in seeming response to the issue of encountering a repeated instruction. In Fig. 16. we see a guest's second attempt at checking-in, where we compare the first and second cycles. Given that this is the second attempt, the guest will have heard the close / open eyes instructions three times already, and here we see how he attempts some change to remedy the lack of success.

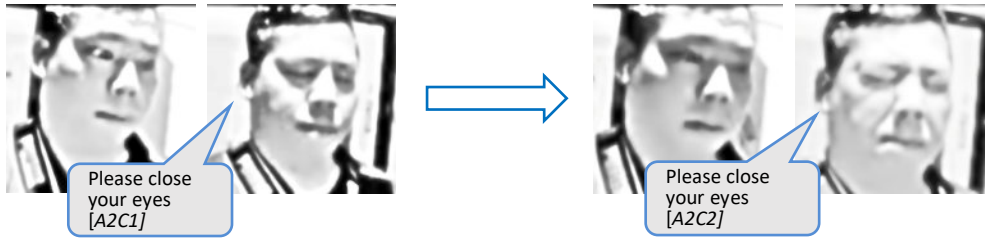

Fig. 16. Guest closing eyes during first cycle (left), then doing so more intensely on next cycle (right)

Similarly to changes in response to the manner in which the eyes were closed, we sometimes observed a change in how guests responded to multiple iterations of the instruction to "shake your head". For example, they might shake their head more quickly, slower, or perform different kinds of left-right movements from first to second try. In Fig. 17. we provide just one example of this change.

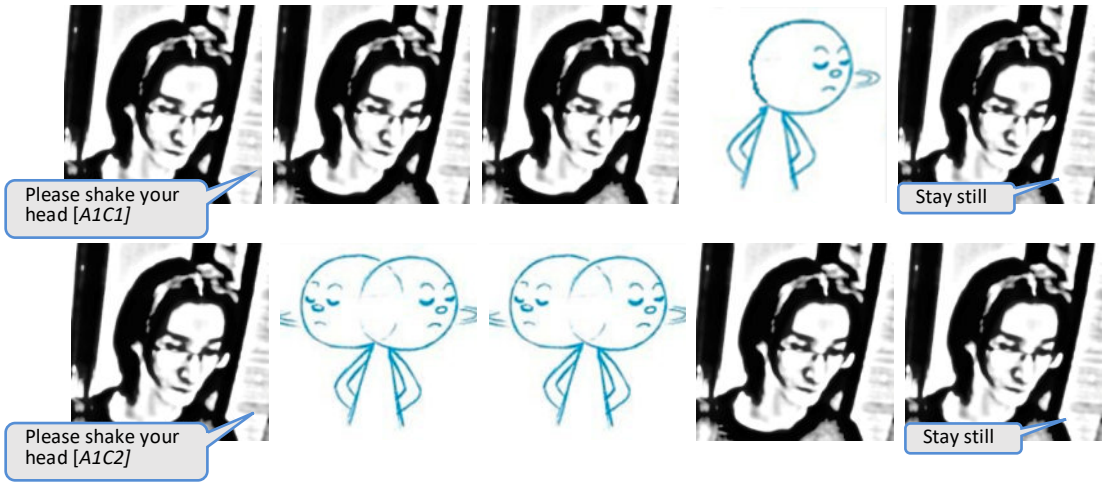

Fig. 17. A guest shaking their head in limited fashion on first cycle (top), doing so more on next cycle (bottom)

7.4 Understanding the paired nature of instructions

We have seen how guests change their appearance, position, and overall response to the kiosk after explicit failures or after noticing they are being issued repeated instructions. However, this is not the only situation that prompts them to do so. Guests also learn that instructions are not 'stand-alone' but actually come in 'pairs'.

If a guest first hears "close your eyes", then this instruction is very 'open': How long should the guest close their eyes? Until *when* should they do so? (see previous section.) Should they just 'blink'? Or should they close their eyes for a long period of time? However, once the guest learns that "close your eyes" is *followed* by the instruction to "open your eyes" from the kiosk, their understanding of the *first* instruction may change by the unfolding revelation of the instruction's paired nature. The guest may conclude that they should close their eyes *until* they hear the instruction "open eyes" next time round. Equally, perhaps this means that all further instructions are also paired. In other words, the unfolding paired nature of the first instruction then configures the guest's response to all subsequent instructions which henceforth may be *heard in pairs*.

Indeed, some guests were faced with an interesting 'problem': in the first cycle, if they had followed the initial instruction, but had done so only for a moment, then their eyes were already

open when they heard the second instruction: "open your eyes". What should they do in that situation? Alternatively, if they had not followed the first instruction ("close your eyes"), their eyes would 'naturally' be open—what should they do then?

Let us look at an example of the former problem. In Fig. 18. we see a guest resolving the 'problem' by opening his eyes wider in response to having 'inadequately' followed the initial instruction. This reaction happened not infrequently. While it seems unlikely that this action was considered within the design process of the facial recognition system and kiosk, it is nevertheless an interesting solution by the guest.

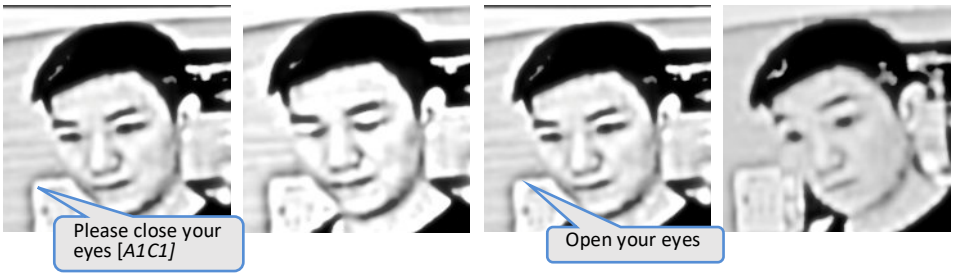

Fig. 18. Guest responding to "open your eyes" when eyes are *already* open

The clearest way to see that guests learn about the paired nature of instructions is to compare their responses to the first time they hear "close your eyes" and the second time they hear "close your eyes", that is, understanding instructions 'inter-cycle'. Many more guests keep their eyes closed *until* they hear "open eyes" in the second cycle. Of the 361 cases in our collection, this happened in 205 cases (57%). In other words, in more than half the cases, the guest had their eyes *open* (on the first cycle of the first attempt) when they heard the second part of the instruction, "open your eyes". Table 1 breaks this phenomenon down further, showing the variety of different responses by guests to resolve this problem.

Table 1: Responses by guests with eyes already open upon hearing the paired instruction "open your eyes"

	Category	**Number Of Cases (Percentage)**
Guests with eyes already open upon hearing "open your eyes" (n = 205)	Do nothing	126 (61%)
	Close & open eyes / blink	32 (16%)
	Open eyes wider	47 (23%)

Based upon the collected instances of guests' actions, we would argue that they learn about the paired nature of instructions *during* the first cycle, i.e., 'intra-cycle'. This conclusion is evidenced in the ways their responses to *subsequent* paired instructions seemed to be framed by their encounter with the first instruction. We can demonstrate this by comparing guest responses to the initial first-pair instruction "close your eyes" with the fifth ("open your mouth", which is followed by "close your mouth").

Table 2: Guests performing two separate, but connected actions: open eyes / close eyes, and open mouth / close mouth

Cases	Sample Size	Guests performing the instructed action until they hear next instruction	Guests stopping their actions before they hear the next instruction
Following the first instruction ("Close your eyes")	n = 298	93 (31%)	205 (69%)
Following the fifth instruction ("Open your mouth")	n = 280	263 (94%)	17 (6%)

Table 2 compares responses by guests to two different instructions. The temporality is important, in that by the time the guest reaches the fifth instruction, on the whole they showed a markedly different orientation: towards "open your mouth" as a prospectively paired instruction. This means that guests overwhelmingly waited for the subsequent instruction to "clos e your mouth", whereas this was far less common for those same guests performing the very first instruction encountered (note the change from 31% to 94%).

8 DISCUSSION

Our analysis has focused on two aspects of the work that is required to make facial recognition work [8, 27]: **preparation work** and **remedial work**. On the one hand, we have shown what people may do *prior* to facial recognition actually being performed. Preparatory work is conducted by guests to curate or 'format' themselves (their appearances and actions) according to what they anticipate as the 'needs' of FRT to ensure success in achieving machine recognition. We found that even bystanders may become entangled in these preparatory activities as computing systems are often operating in public or semi-public environments [46, 47]—in this sense we must shift our conceptualizations of face recognition technologies to appreciate how they operate in socially organized settings in which members may be closely coordinating actions oriented to the fact of FRT in place. More simply, we see how (interactive) FRT relies on prospective enactments of compliance from people *proximate* to them, regardless of the designers' expectations of who 'the user' actually is.

On the other hand, we have demonstrated how troubles noticed *during* the process of interactive facial recognition leads users to continuously make changes in their appearance and manage their performance of particular kinds of facial and bodily actions and comportments. We called this remedial work. This sometimes involves very small changes made by the people being recognized, within the bounds of their attempts at producing compliant responses to instructions, as well as how those instructions were understood in sequence and as pairs. Once again, bystanders also came into play as further sources of disruption for being facially recognized. The sum total of this work is what this particular facial recognition technology entails for those interacting with it.

In the rest of this discussion we want to reflect on what this means for studies of facial recognition in CSCW by making five points: first, that facial recognition is not 'automatic' but requires work on the part of guests; second, that FRT usability and accessibility based on actual instances 'in the wild' need addressing in terms of socially organized, collaborative work they may entail; third, that people work to 'discipline' themselves [58, 59] according to the (perceived) machinic requirements; fourth, that the relationship of an interactive FRT system has with the *situation* has consequences for acceptability; and fifth, that CSCW (and HCI more broadly) can learn new things about facial recognition technologies in practice and develop new concepts like interactional compliance, or disciplining work, which enrich and mature the approach in research on FRT.

Our first point is simply that **facial recognition is not automatic; it takes work**. This conclusion seems obvious in hindsight, having examined the video recordings from our study, but this observation rarely seems to factor into discussions of facial recognition technologies. This lack of attention might be because the concept of recognition strongly suggests that it is an *automatic* process applied to people. In the case of the guests checking-in to the hotel, facial recognition is used to 'automatically' verify someone's identity for the purposes of matching them up to a booking and associated ID card. Like many facial recognition systems used for authentication, it is often argued that applying facial recognition improves usability and simplicity (no need to remember anything, just 'bring your face', etc.). However, as we have shown, now guests as customers really just have to do *more* work to be recognized: in other words they have become the 'working customer' [25, 48, 49]. The nature of this work tends to be forgotten in the discourse surrounding AI systems. Designers want facial recognition technologies to be 100% accurate, whereas critics point out that there can easily be misidentification [12]; however, flaws in recognition are countered with the idea that gradual improvements will remove these flaws. This proposition seems slightly fanciful, however, because of what we think underlies all of this: a confusion between machine recognition, and what we conventionally consider when we recognize someone. 'Recognition' suggests capabilities that are simply not present in the machine; it is not only that (as others have pointed out) FRTs get matches wrong frequently in non-trivial applications, but, as we have shown, getting recognized can involve a not insignificant amount of facial and bodily work, as well as social coordination on the part of the person and those that are nearby. On this basis, we can hardly think that this is suitably spoken of as recognition in any conventional sense and to suggest otherwise is deceptive. Further, machine 'recognition', must be understood as more than just a dyadic human-machine phenomenon.

Second, our study highlights some thorny problems with **the usability and accessibility of interactive facial recognition technologies** in circumstances like identification. Although most guests managed to check-in successfully, a paucity of feedback from the kiosk hindered what guests could do to resolve problems, meaning that they relied both upon background knowledge about what interactional work to perform, and something akin to trial and error in dealing with any trouble they encountered. Further, given the 'black box' nature of facial recognition systems, it is unclear how much comprehensible feedback could be easily designed. It is possible that the 'explainable AI' movement [37] may present some ways of enabling greater interactional purchase for the FRT user in these cases, but such advances notwithstanding, inscrutability remains a significant limitation of enabling basic usability improvements that render FRT more socially accountable to a given situation. In that sense, it may be that designers are limited in what they can do here. One possible resolution is to look to service workers' body of knowledge in dealing with common problems and ways to resolve trouble.

Connected with usability is accessibility. We note that the facial recognition system here—like other recognition technologies—relies upon certain kinds of facial (and bodily) comportments or actions that are assumed to be producible by 'any' guest. Here we can see how the work of guests interacting with the kiosk turns upon body and face, and this necessary compliance with the facial recognition system connects that compliance with specific assumed abilities, thus potentially excluding people with disabilities that, say, prevent the production of said comportments or actions. This problem of bodily normativity has been noted elsewhere recently [55] and also relates to the categorization issues we mention next. It also follows that—although our study has not examined this—interactive FRT like the system we study, has the potential to perpetuate harms previously discussed, like failing to detect faces based on gender, skin tone, or other characteristics.

Our third point is that in doing the practical work of making themselves recognizable, **people discipline themselves to the anticipated requirements of machinic recognition, no matter whether they are directly interacting or bystanding**. This result seems complementary to prior observations about FRT embedding particular kinds of categorizational

frames (e.g., of race, gender, emotion, age, etc.) [53, 56], except we broaden this to consider the socially situated nature of FRT. Our analysis has shown how guests and bystanders transform their bodies and faces into an appropriately 'formatted input' for facial recognition systems. There is no softness or pliability to the requirements of compliance. This kind of disciplining is 'hard' in that the conditions set by the facial recognition technologies must be met by its users and those that are bystanding. No real help is provided to guide this disciplining work; hence we uncover significant self-remedial work, which is reminiscent of trial-and-error 'debugging'. This work contrasts to other settings where compliance is achieved through 'narrowing' the *instruction* (rather than changing the *response*). For example, in classrooms teachers often modify their initial question after a pupil's inadequate response, following a 'funnel pattern' [62]. The overlooked centrality of guests (and bystanders) needing to discipline their behavior—bodily, facial, verbal or otherwise—is, we believe, derived from a conceptual mistake about recognition. We would argue that the very idea of machine recognition always relies on human self-discipline of some sort or another, quite opposite to the promise of machine recognition and automation.

Our observations are also related to Foucault's notions of 'self-discipline' and 'technologies of self' [18, 19, 41], "which permit individuals to effect by their own means, or with the help of others, a certain number of operations on their own bodies and souls, thoughts, conduct, and way of being, so as to transform themselves in order to attain a certain state of happiness, purity, wisdom, perfection, or immortality" [19, p. 225]. However, while Foucault's focus is on the way people's transformations are driven by what society deems desirable (see [20]), here we have a transformation that the *machine* 'deems' desirable (or more properly, that the designers or engineers have—explicitly or not—deemed desirable). In other words, the 'self-discipline' that we observe in these instances means the guests in particular are thoroughly subjugating to the machine.

This subjugation to the machine becomes particularly visible when guests are not alone but accompanied by (bystanding) roommates. It is in these situations that the objectifying and embarrassing aspects of the process becomes in the reactions of bystanders. For instance, in Fig. 19. we see, (a) a bystanding roommate smiling at the guest, and (b) the bystander jokingly disrupting the bodily position of the guest with a gentle 'kick'.

(a) (b)

Fig. 19. Bystanding roommates reacting to what guests are doing

As Barthes says of the experience of being photographed, "I am ... a subject who feels he is becoming an object" [3]. We see the examples highlighted in Fig. 19. very much echoing the "object[ification]" described by Barthes. The need for disciplining can thus *sometimes* also be dehumanizing too, requiring a performance of sorts, oftentimes a reluctant one that exposes oneself to ridicule. This sense is quite the opposite of suggestions in HCI that AI technologies may be potentially 'empowering' users [54], indicating some potential design sensitivities that need further care. However, this is *not* to say that dehumanization is somehow a deterministic end product of disciplining. We do not intend to equate the need for self-disciplining with dehumanization. For instance, disciplining work is crucial to both gamers and musicians (to pick two examples) in their practices: it underpins expertise, accomplishment, and artistic expression as well as collaboration in the form of audience-performer interactions. The point, then, is that

disciplining is a broad phenomenon that needs greater scrutiny with respect to AI technologies purporting to deliver recognition; there is a need to take stock of what disciplining leads to, which may, sometimes, be dehumanization.

Our fourth point is about ***when* to use facial recognition**. Although not without its issues [56], facial recognition employed for identity verification like Apple's Face ID seems less problematic because the user is already used to typing in a code, using a fingerprint, or providing another means of authentication. In other words, you are replacing one interaction with a machine for another. Passport control is another situation where facial recognition is used for biometric ID. Although we are replacing an interaction with a human border control officer with a human-machine interaction both situations seem to be by default very personal and invasive albeit in different ways. Yet the use of facial recognition in hotels for checking-in or in retail for payment is very different from these other examples: facial recognition technologies replaces (or at least tries to replace) a service worker's role with guests. Verification at a hotel used to be based on a *human, social encounter*, with all that this coordinatively entails (small talk, care, etc.). Now it is a primarily *machine encounter* with occasional interventions by others (bystanders, hotel staff, etc.). We made the point previously that service worker knowledge is also crucial to ensuring that such technologies as facial recognition technologies work successfully, practices that are often ignored by formal accounts of work, leading to deskilling [44]. While facial recognition technologies in these cases is sold as 'convenience', what we show is that it actually just means more work done by the guest (and roommates).

For the fifth and final point, both work (our 1st point) and 'self-discipline' (3rd point) has implications for **CSCW's approach to studies of facial recognition technologies**. Primarily studies have tended to focus on facial recognition's use in surveillance of one kind or another (formal or otherwise), with good reason given the ways in which facial recognition technologies have been increasingly deployed in decision-making structures of consequence (policing, the state, workplaces, etc.) and with inbuilt assumptions about what the face signifies. But rarely (to the best of our knowledge) have critical examinations of facial recognition technologies looked at the concrete circumstances of their use 'on the ground' and in detail within social settings; in this regard we think that concepts like **interactional compliance** and **disciplining work** can be added to CSCW's conceptual framework for how to understand and talk about some implementations of facial recognition in society. Thus, to close, we think this paper may act as a pointer towards broadening this program methodologically, providing a further check on the inflated expectations that run alongside deployments of AI technology seemingly without end.

9 CONCLUSION

The study we have presented here has offered a first look at how people encounter an interactive facial recognition system. Although the system itself—a hotel check-in—is, at first blush, seemingly less consequential than more concerning uses of facial recognition, for example, surveillance—perhaps one might even say interactive facial recognition here is 'gimmicky' (featured as part of a 'robotic hotel' chain)—we nevertheless find much richness here to help develop our understanding of what it means to be the object of machine recognition, and what the nature of this concept of recognition truly entails as a practical matter. Deploying facial recognition technologies, we find, means careful disciplining work on the part of guests and bystanders, so as to produce compliance with the needs of the machines (and their design).

This paper is obviously limited to one particular instance of facial recognition and its interaction design as implemented in the check-in kiosk system and interface. However, the concepts of self-disciplining work, its underexamined role in human practices with and around recognition technologies like FRT, and the potential for subjugation or dehumanization of 'users'—these remain significant topics for future investigation of a much wider range of AI-driven interactive technologies that are being deployed in more and more mundane circumstances. To wit, we encourage others to take seriously the value of close examinations of

practical mundanity to fill out our understanding of what it means when AI technologies like facial recognition technologies are placed 'in the world'.

ACKNOWLEDGEMENTS

This work would not be possible without the generosity of our participants. We are incredibly grateful to all the customers who participated in the research as well as the service workers and managers who helped with the fieldwork and data collection. This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/T022493/1] and the Royal Society Kan Tong Po International Fellowship [grant number KTP\R1\211015].

REFERENCES

1 Anton Alterman. 2003. 'A piece of yourself': Ethical issues in biometric identification. *Ethics and Information Technology* 5, 3 (2003), 139-150. DOI: https://doi.org/10.1023/B:ETIN.0000006918.22060.1f

2 Fabio Bacchini and Ludovica Lorusso. 2019. Race, again: how face recognition technology reinforces racial discrimination. *Journal of Information, Communication and Ethics in Society* 17, 3 (August 2019), 321–335. DOI: https://doi.org/10.1108/JICES-05-2018-0050

3 Justinas Baltrusaitis. 2021. Top 10 Countries and Cities by Number of CCTV Cameras. Precise Security (26 August 2021). Available at https://www.precisesecurity.com/articles/top-10-countries-by-number-of-cctv-cameras (Accessed 6 January 2022).

4 Roland Barthes. 1982. *Camera Lucida: Reflections on Photography* (Second ed.). Hill and Wang, New York, NY.

5 Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. 2019. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest* 20, 1 (July 2019), 1–68. DOI: https://doi.org/10.1177/1529100619832930

6 Cynthia L. Bennett and Os Keyes. 2020. What is the point of fairness? Disability, AI and the complexity of justice. *ACM SIGACCESS Accessibility and Computing* 125 (March 2020), Article No. 5. DOI: https://doi.org/10.1145/3386296.3386301

7 Jeremy Bentham. 1843. The principles of penal law. In *The Works of Jeremy Bentham: Volume* 1 William Tait, Edinburgh, 367-380.

8 Masha Borak. 2021. Facial recognition is used in China for everything from refuse collection to toilet roll dispensers and its citizens are growing increasingly alarmed, survey shows. *South China Morning Post* (27 January 2021). Available at https://www.scmp.com/tech/innovation/article/3119281/facial-recognition-used-china-everything-refuse-collection-toilet (Accessed 6 January 2022).

9 John Bowers. 1994. The work to make a network work: Studying CSCW in action. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work (CSCW '94)*, ACM, New York, NY, 287–298. DOI: https://doi.org/10.1145/192844.193030

10 Kieran Browne, Ben Swift, and Terhi Nurmikko-Fuller. 2020. Camera adversaria. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, ACM, New York, NY. DOI: https://doi.org/10.1145/3313831.3376434

11 Claudio Celis Bueno. 2020. The face revisited: Using Deleuze and Guattari to explore the politics of algorithmic face recognition. *Theory, Culture & Society* 37, 1 (January 2020), 73–91. DOI: https://doi.org/10.1177/0263276419867752

12 Kate Crawford. 2019. Regulate facial-recognition technology. *Nature* 572, 7771 (August 2019), 565–565. DOI: https://doi.org/10.1038/d41586-019-02514-7

13 Jose Sanchez del Rio, Daniela Moctezuma, Cristina Conde, Isaac Martin de Diego, and Enrique Cabello. 2016. Automated border control e-gates and facial recognition systems. *Computers & Security* 62, (September 2016), 49–72. DOI: https://doi.org/10.1016/j.cose.2016.07.001

14 Michel De Fornel. 1996. The interactional frame of videophonic exchange. *Réseaux* 4, 1 (1996), 47–72. DOI: https://doi.org/10.3406/reso.1996.3305

15 Alexander De Luca, Alina Hang, Emanuel von Zezschwitz, and Heinrich Hussmann. 2015. I feel like I'm taking selfies all day! Towards understanding biometric authentication on smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, ACM, New York, NY, 1411–1414. DOI: https://doi.org/10.1145/2702123.2702141

16 Paul Dourish, Annette Adler, Victoria Bellotti, and Austin Henderson. 1996. Your place or mine? Learning from long-term use of Audio-Video communication. *Computer Supported Cooperative Work* 5, 1 (March 1996), 33–62. DOI: https://doi.org/10.1007/BF00141935

17 Anna Filipi. 2009. Toddler and Parent Interaction: The Organisation of Gaze, Pointing and Vocalisation. John

Benjamins, Amsterdam.

18 Michel Foucault. 1977. Discipline and Punish: The Birth of the Prison. Allen Lane, London.

19 Michel Foucault. 1997 [1982]. Technologies of the self. In *Ethics: Subjectivity and Truth*. New Press, New York, NY, 223-251.

20 Arthur W. Frank, and Therese Jones. 2003. Bioethics and the later Foucault. *Journal of Medical Humanities*, 24, 3/4 (Winter 2003), 179-186. DOI: https://doi.org/10.1023/A:1026014718326

21 Hannah Fry. 2018. Hello World: How to be Human in the Age of the Machine. W. W. Norton, New York, NY.

22 Yumei Gan, Christian Greiffenhagen, and Stuart Reeves. 2020. Connecting distributed families: Camera work for three-party mobile video calls. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, ACM, New York, NY. DOI: https://doi.org/10.1145/3313831.3376704

23 Kelly A. Gates. 2011. Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance, New York University Press, New York, NY.

24 Timit Gebru. 2018. Understanding the limitations of AI: When algorithms fail. https://smartech.gatech.edu/handle/1853/60425 (Accessed 6 January 2022).

25 Cathy Goodwin. 1988. 'I can do it myself': Training the service consumer to contribute to service productivity. *Journal of Services Marketing* 2, 4 (April 1988), 71–78. DOI: https://doi.org/10.1108/eb024745

26 Olga Goriunova. 2019. The digital subject: People as data as persons. *Theory, Culture & Society* 36, 6 (November 2019), 125–145. DOI: https://doi.org/10.1177/0263276419840409

27 Rebecca E. Grinter, W. Keith Edwards, Mark W. Newman, and Nicolas Ducheneaut. 2005. The work to make a home network work. In *Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work (ECSCW '05)*, Springer, Berlin, 469–488. DOI: https://doi.org/10.1007/1-4020-4023-7_24

28 Adam Harvey. 2020. Computer Vision Dazzle Camouflage. https://cvdazzle.com (Accessed 6 January 2022).

29 Christian Heath and Paul Luff. 1991. Collaborative activity and technological design: Task coordination in London underground control rooms. In *Proceedings of the Second European Conference on Computer-Supported Cooperative Work (ECSCW '91)*. Springer, Dordrecht, 65–80. DOI: https://doi.org/10.1007/978-94-011-3506-1_5

30 Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in Qualitative Research*. Sage Publications.

31 Alexa Hepburn and Galina B. Bolden. 2013. The conversation analytic approach to transcription. In *The Handbook of Conversation Analysis*. Blackwell, Oxford, 57-76.

32 Kashmir Hill. 2020. Wrongfully accused by an algorithm. *The New York Times* (24 June 2020). Available at https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html (Accessed 6 January 2022).

33 Yarden Katz. 2020. Artificial Whiteness, Politics and Ideology in Artificial Intelligence. Columbia University Press, New York, NY.

34 Os Keyes. 2018. The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on Human-Computer Interaction, 2 (CSCW)*, 1–22. DOI: https://doi.org/10.1145/3274357

35 Genia Kostka, Léa Steinacker, and Miriam Meckel. 2021. Between security and convenience: Facial recognition technology in the eyes of citizens in China, Germany, the United Kingdom, and the United States. *Public Understanding of Science* 30, 6 (August 2021), 671–690. DOI: https://doi.org/10.1177/09636625211001555

36 Christian Licoppe and Julien Morel. 2009. The collaborative work of producing meaningful shots in mobile video telephony. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09)*. ACM, New York, NY. DOI: https://doi.org/10.1145/1613858.1613903

37 Zachary C. Lipton. 2018. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16, 3 (May-June 2018), 31–57. DOI: https://doi.org/10.1145/3236386.3241340

38 Georgina Lukanova and Galina Ilieva. 2019. Robots, artificial intelligence, and service automation in hotels. In *Robots, Artificial Intelligence, and Service Automation in Travel, Tourism and Hospitality*. Emerald, Bingley, 157–183. DOI: https://doi.org/10.1108/978-1-78756-687-320191009

39 David Lyon. 2001. *Surveillance Society: Monitoring Everyday Life*. Open University Press, Buckingham, UK.

40 Douglas Macbeth. 2004. The relevance of repair for classroom correction. *Language in Society* 33, 5 (November 2004), 703–736. DOI: https://doi.org/10.1017/S0047404504045038

41 Ivan Manokha. 2018. Surveillance, panopticism, and self-discipline in the digital age. *Surveillance & Society* 16, 2 (July 2018), 219–237. DOI: https://doi.org/10.24908/ss.v16i2.8346

42 Robert J. Moore. 2008. When names fail: Referential practice in face-to-face service encounters. *Language in Society* 37, 3 (July 2008), 385–413. DOI: https://doi.org/10.1017/S004740450808055X

43 Robert J. Moore, Jack Whalen, and E. Cabell Hankinson Gathman. 2010. The work of the work order: Document practice in face-to-face service encounters. In *Organisation, Interaction and Practice Studies of Ethnomethodology and Conversation Analysis*, Cambridge University Press, Cambridge, 172-197.

44 Julian E. Orr. 1996. *Talking About Machines: An Ethnography of a Modern Job*. Cornell University Press, Ithaca, NY.

45 Daniella Raz, Corinne Bintz, Vivian Guetler, Aaron Tam, Michael Katell, Dharma Dailey, Bernease Herman, P. M. Krafft, and Meg Young. 2021. Face mis-ID: An interactive pedagogical tool demonstrating disparate accuracy rates in facial recognition. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (AIES-2021), ACM, 895–904. DOI: https://doi.org/10.1145/3461702.3462627

46 Stuart Reeves. 2011. *Designing Interfaces in Public Settings*. Springer, London. DOI: https://doi.org/10.1007/978-0-85729-265-0

47 Stuart Reeves, Mike Fraser, Holger Schnädelbach, Claire O'Malley, and Steve Benford. 2005. Engaging augmented reality in public places. In *Adjunct Proceedings of 2005 SIGCHI Conference on Human Factors in Computing Systems* (*CHI 2005*).

48 Kerstin Rieder, and G. Günter Voß. 2010. The working customer: An emerging new type of consumer. *Psychology of Everyday Activity*, 3, 2 (2010), 2-10.

49 Kerstin Rieder and G. Günter Voβ. 2013. The working customer: A fundamental change in service work. In *Customers at Work*, Palgrave Macmillan, London, 177–196. DOI: https://doi.org/10.1057/9781137293251_9

50 Craig Robertson. 2009. A documentary regime of verification: The emergence of the US passport and the archival problematization of identity. *Cultural Studies* 23, 3 (May 2009), 329–354. DOI: https://doi.org/10.1080/09502380802016253

51 Robotics Business Review (2018). Facial Recognition Cameras and AI: 5 Countries With the Fastest Adoption. Robotics Business Review (5 December, 2018). https://www.roboticsbusinessreview.com/ai/facial-recognition-cameras-5-countries/ (Accessed 6 January 2022).

52 Morgan Klaus Scheuerman, Jacob M. Paul, and Jed R. Brubaker. 2019. How computers see gender: An evaluation of gender classification in commercial facial analysis services. In *Proceedings of the ACM on Human-Computer Interaction*, 3 (CSCW), 1–33. DOI: https://doi.org/10.1145/3359246

53 Morgan Klaus Scheuerman, Kandrea Wade, Caitlin Lustig, and Jed R. Brubaker. 2020. How we've taught algorithms to see identity: Constructing race and gender in image databases for facial analysis. In *Proceedings of the ACM on Human-Computer Interaction,* 4 (CSCW), 1–35. DOI: https://doi.org/10.1145/3392866

54 Ben Shneiderman. 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (April 2020), 495–504. DOI: https://doi.org/10.1080/10447318.2020.1741118

55 Katta Spiel. 2021. The bodies of TEI: Investigating norms and assumptions in the design of embodied interaction. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '21)*, Article No. 32. ACM. DOI: https://doi.org/10.1145/3430524.3440651

56 Luke Stark. 2019. Facial recognition is the plutonium of AI. *XRDS* 25, 3 (April 2019), 50–55. DOI: https://doi.org/10.1145/3313129

57 Lucy A. Suchman. 1987. Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge University Press, Cambridge.

58 Burak Tekin. 2019. Bodies at Play: Exploring Participation, Spectatorship, and Morality in Videogaming Activities. PhD thesis, University of Basel, Basel, Switzerland.

59 Burak S. Tekin and Stuart Reeves. 2017. Ways of spectating: Unravelling spectator participation in Kinect play. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17).* ACM, New York, NY, 1558–1570. DOI: https://doi.org/10.1145/3025453.3025813

60 Gabriel H. Teninbaum. 2021. Report on ExamSoft's ExamID feature (and a method to bypass It). Forthcoming in *Journal of Robotics, Artificial Intelligence, and* Law. DOI: https://doi.org/10.2139/ssrn.3759931

61 Shui-Hua Wang, Preetha Phillips, Zheng-Chao Dong, and Yu-Dong Zhang. 2018. Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm. *Neurocomputing* 272, (January 2018), 668–676. DOI: https://doi.org/10.1016/j.neucom.2017.08.015

62 Terry Wood. 1994. Patterns of interaction and the culture of mathematics classrooms. In *Cultural Perspectives on the Mathematics Classroom*, Springer, Dordrecht, 149-168. DOI: https://doi.org/10.1007/978-94-017-1199-9_10

63 Xinhua. 2021. 人脸识别技术推广受信息泄露困扰 [Facial recognition technology promotion is plagued by information leakage]. Xinhua (7 April 2021). Available at http://www.xinhuanet.com/2021-04/07/c_1127300685.htm (Accessed 6 January 2022).

64 Xinhuanet. 2021. China's top court specifies rules for facial recognition. Xinhuanet (28 July 2021). Available at http://www.xinhuanet.com/english/2021-07/28/c_1310092626.htm (Accessed 6 January 2022).

65 Yutaka Yamauchi, Jack Whalen, and Daniel G. Bobrow. 2003. Information use of service technicians in difficult cases. In *Proceedings of the Conference on Human Factors in Computing Systems* (*CHI '03).* ACM, New York, NY, 81-88. DOI: https://doi.org/10.1145/642611.642627

66 Phoebe Zhang. 2020. Privacy in China: the growth of facial recognition technology in the private sector raises concerns about security and identity. *South China Morning Post* (26 November 2020). Available at https://www.scmp.com/lifestyle/article/3111428/privacy-china-growth-facial-recognition-technology-private-sector-raises (Accessed 6 January 2022).