AI Emotion Recognition and Affective Injustice

Michael Dale and Steven Gubka[1]

Penultimate draft — Please cite final version forthcoming in *Erkenntnis*

Abstract: Artificial intelligence can now recognize our emotions using algorithms that interpret our facial expressions. This technology is used to help assess an applicant's interview performance, an individual's potential for criminal behavior, whether a student is paying attention during an online class, and more. Assuming that such technology could reliably recognize human emotions, it nonetheless cannot assess whether an emotion is apt, which matters for how we ought to treat someone. Specifically, we argue that such uses of AI Emotion Recognition constitute *affective injustice* that occurs when someone's emotions are treated unjustly. We hope to draw attention to this issue so that designers and proponents of AI Emotion Recognition recognize the principled limitations of the technology in its current state.

Keywords: artificial intelligence; emotion recognition; affective injustice; affective indiscrimination

1. Introduction

Artificial intelligence can now recognize our emotions using algorithms that interpret our facial expressions (e.g. Affectiva, 2023). This technology is used to help assess an applicant's interview performance, the potential for criminal behavior such as shoplifting in a store, whether a student is paying attention during an online class, and more (Kaye, 2022; Martinez-Martin, 2019; Thomas, 2018; Zetlin, 2018; Zhao and Tsai, 2011).

Two challenges have been raised to the accuracy of the technology. The first is that facial expressions are insufficient as evidence of emotions because there is no reliable connection between facial expressions (such as a scowl) and emotions (such as anger) (Barrett et al., 2019; Duran et al., 2017; Fernández-Dols, 2017). The second is that the data sets used to 'train' AI to interpret facial expressions as expressing a particular emotion are biased against racial and cultural minorities (Buolamwini & Gebru, 2018; Kim et al., 2021; Rhue, 2018). We examine

[1] Both authors contributed equally to this manuscript.

these challenges and note how potential advancements in the technology may overcome these limitations, such as using more diverse data sets.

AI Emotion Recognition faces another significant burden in its use cases that has been unexplored. The technology cannot assess whether an emotion is apt, which matters for how we ought to treat someone. For example: an applicant's anger may be apt if the interviewer has performed a microaggression against them. But without an assessment of the applicant's emotion as apt or not, the fact that they feel angry is not yet a reason to rank them as a worse applicant. Of course, AI Emotion Recognition does not reveal anything about whether someone's emotion is apt (or even what that emotion is directed toward).

This suggests that, even if AI Emotion Recognition can reliably recognize human emotions, it may be used carelessly if its limitations with respect to the aptness of emotions are not understood. Specifically, we raise the worry that such uses of AI Emotion Recognition facilitate *affective injustice* that occurs when "the treatment of emotions is unjust, or emotions are used to treat people unjustly" (Pismenny et al., 2024: 154). One such type of wrongdoing is to pressure someone to cease feeling or expressing an apt emotion (Srinivasan, 2018). Along similar lines, we argue that AI Emotion Recognition can facilitate *affective indiscrimination*, which we characterize as an incorrect doxastic attitude toward someone's apt emotions. Although AI Emotion Recognition is not inherently unjust, one type of misuse of the technology involves the disrespect of apt emotions.

Here is the plan. In the next section, we explain the recent challenges to the accuracy of AI Emotion Recognition and briefly discuss how those challenges might be met in the future. In section three, we develop the main claim of the paper that misuse of AI Emotion Recognition can result in affective injustice of a specific type that we call affective indiscrimination, and in

section four, we defend this claim from objections. Section five concludes with some prospects for avoiding affective injustice in the use and design of AI Emotion Recognition.

2. Accuracy Challenges for Emotion Recognition Technology

Two of the most salient criticisms of emotion recognition technology concern the accuracy of the technology. In this section, we describe these criticisms, and then argue that even though they are significant problems for the current state of the art in emotional recognition technology, they can be addressed with more developed and nuanced technology. Which is to say, they are not reasons to stop emotion recognition research; they are reasons to improve it.

2.1. Facial Expression and Emotion

It is commonly believed that particular facial expressions reliably signal certain emotional categories in such a way that when someone has one of these facial expressions, it can be inferred that they are experiencing a certain emotional state. For example, a scowl is thought to reliably signal the emotional state of anger. This *common view* (Barrett, 2017; Barrett et al., 2019) seems to be widespread.[2] Indeed, children are often taught about emotion by being shown pictures and videos of certain emotional expressions (think of Sesame Street and Miniland Educational Games), and most forms of media that use photos and video take it for granted that certain expressions reliably signal particular emotional states (e.g., if Jim Carrey is seen smiling, it is commonly understood that he is feeling the emotional state of happiness). We even send and receive caricatures of these emotional expressions (known as emojis or emoticons) to communicate our emotional states.[3]

[2] To be clear, the common view should not be understood as a particular view held in the literature. Instead, it is best described as a collection of assumptions that many people hold regarding the relationship between facial expressions and emotional categories. For a full articulation of the common view, see Barrett (2017) and Barrett et al. (2019).

[3] The common view also reveals itself in more professional settings. Legal actors across the Western world (e.g., judges and jurors) often "read" the facial expressions of defendants to help determine their guilt (Bandes, 2014). Moreover, the Reading the Mind in the Eyes Test–which asks test takers to match facial expressions with particular

Many technology companies are developing AI Emotion Recognition technology in accordance with this common view. For example, Microsoft's Emotion API claimed to be able infer people's emotional states based on a certain "reading" of their expressions in photos and videos.[4] Similarly, Zoom has been developing a program that supposedly analyzes the facial expressions of zoom video callers in such a way that it can then give hosts assessments of the emotional states of the people in the meeting (Kaye, 2022). Furthermore, UK firm WeSee uses AI Emotion Recognition technology to assess whether someone's underlying emotions during interviews with law enforcement indicate if they are a potential threat (Thomas, 2018). And in the realm of education, Century Tech employs an AI system to assess students' emotional responses to a lesson, and then adjusts the lesson difficulty accordingly (PYMNTS, 2024).

However, these AI programs have been met with criticism, due to an increasing amount of evidence in social and developmental psychology that calls into question the veracity of the common view. In a meta-analysis of studies in which participant's expressions were analyzed when they were faced with an emotion inducing stimulus, it was found that participants exhibited a range of expressions in response to each emotional stimulus, only some of which matched the common view (Duran et al., 2017). Studies looking at spontaneous expressions in more naturalistic settings (e.g., when an athlete was talking with an audience) found similar results (Fernández-Dols, 2017). To take just one example, a Duchenne smile, which is commonly thought to be an expression of happiness, was also found to be spontaneously produced when

emotion words–is widely utilized in autism research (Baron-Cohen et al., 2001), and learning to recognize these facial configurations as emotional expressions is often part of treatment plans for autism and other psychiatric disorders (Baron-Cohen et al., 2004).

[4] In response to the kind of criticism we discuss in this section, Microsoft has recently ended this program (Hill, 2022).

people were not happy (Gunnery & Hall, 2014), likely signaling submission or affiliation instead (Rychlowska et al., 2017).

Moreover, there is even some (albeit limited) evidence in cross-cultural studies that suggests that people in small-scale remote cultures exhibit facial expressions not in line with the common view. For example, Paul Ekman (1980) took photographs of the facial expressions of Fore hunter-gatherers in Papua New Guinea, and it was later found out that Trobriand Islanders of Papua New Guinea and the Maori of New Zealand interpreted these facial expressions differently than westerners (Barrett et al., 2019; Crivelli et al., 2017). In one case, the expression that westerners described as a fear expression was interpreted as an expression of threat and anger by the Islanders and the Maori. This suggests that there is more inter-cultural variation for facial expressions of emotions than the common view traditionally allows for.

Due to this (and other) evidence, Barrett et al. (2019) contend that there are three shortcomings in the common view research program. First, it has "*limited reliability* (i.e., instances of the same emotion category are neither reliably expressed through nor perceived from a common set of facial movements)." Second, it has a "*lack of specificity* (i.e., there is no unique mapping between a configuration of facial movements and instances of an emotion category)." Third, it has "*limited generalizability* (i.e., the effects of context and culture have not been sufficiently documented and accounted for)" (Barrett et al., 2019: 3-4). This evidence suggests that any technology company that builds its programs in accordance with the common view is under a mistaken assumption that could lead to widespread inaccuracies.

2.2. Race, Gender, and Age Bias

Empirical research has revealed that facial recognition technology has higher error rates for certain demographics. For example, darker-skinned females are more likely to be

misclassified than fairer-skinned males by commercially available gender classification software (Buolamwini & Gebru, 2018; Kim et al., 2021). Similar effects have been found in emotion recognition technology. Rhue (2018) found that Microsoft's Face++ interpreted the expressions of black basketball players as angrier than white basketball players, even when controlling for the degree to which they were smiling. It was also more likely to interpret a black player's ambiguous expression as contemptuous. Moreover, Kyriakou et al. (2020) found that Microsoft's emotion recognition technology was more likely to correctly infer an angry expression on a black man than a white man, and Amazon's was more likely to incorrectly interpret happy black women as angry than happy white women.

Age has also been shown to be a predictive factor. Kim et al. (2021) looked at emotion recognition accuracy for Amazon Rekognition, Face++, Microsoft Face, and Sightsound, and they found that, overall, these systems performed better on younger faces than older ones. Indeed, the four algorithms could not get above a mean accuracy of 68% in emotion recognition for older adults, whereas mean accuracy for younger and middle aged adults was often above 85%. The effect was especially strong for anger.

These studies suggest that the state of the art in emotion recognition technology exhibits biases in gender, race, and age in line with already existing social stereotypes. This is a significant problem. Not only will this lead to certain demographic groups less inclined to use certain facial recognition software (which may put them at a disadvantage in certain situations in which using the software, or at least being familiar with the software, would be beneficial); it will also amplify stereotypes that are harmful to historically disadvantaged populations.

2.3. Overcoming the Accuracy Challenges

While we grant that these challenges are important to consider, and that they have not yet been adequately addressed, we don't believe that they are reasons to cease research in emotion recognition technology. Indeed, we believe that they can be overcome, with improved technological sophistication. For instance, it seems likely that larger and more diverse data sets will go a long way in addressing at least some of the criticisms about stereotyping. Klare et al. (2012) found that training face recognition software on particular race and gender cohorts improved the accuracy of the software. This is hopeful news, and it suggests that additional technological sophistication and training will help address the accuracy challenges.

Furthermore, it will be important for emotion recognition technology to have a more complex understanding of situations. For example, when an emotion recognition program is being used to assess the emotional state of a defendant on trial, the program should not only pay attention to the facial expressions of the interviewee. This assumes the common view about the connection of facial expressions and emotional states, and, as we just saw, there is good reason to doubt the veracity of this outlook. Instead, the program should take into consideration multiple factors, such as the defendant's bodily gestures, how they are expressing themselves, how they are interacting with those around them, how previous defendants have acted in similar situations, and even small details like the temperature in the room. This kind of technological sophistication will help emotion recognition technology have a more nuanced understanding of the people it is assessing as well as the situations they are in.

Of course, we don't pretend to have made a complete case against these two important accuracy challenges, and we acknowledge that such technology still has a ways to go before adequately addressing them. However, we do believe that addressing these accuracy challenges

is *possible*, and we have suggested some ways in which they may, at some point in the future, be overcome.

3. Emotion Recognition Technology and Affective Injustice

With that said, there's a more significant worry to consider. Suppose that the accuracy challenges can be satisfactorily addressed as we suggested above, and AI Emotion Recognition can reliably categorize our emotions on the basis of our facial expressions. Nonetheless, we argue that it is morally objectionable to use the information from AI Emotion Recognition as though it is evidence about whether someone's emotion is *apt* (that is, whether that emotion is a correct response to the object of that emotion in the way that we will explain below). This is because the fact that someone is feeling an emotion does not (by itself) provide any evidence about whether that emotion is apt. But the mistake here is not merely believing without sufficient evidence: we claim that to falsely believe that someone's emotion is inapt is a case of affective injustice that we call affective indiscrimination.

3.1. The Misuse of AI Emotion Recognition in Interviews

Let us illustrate this challenge for AI Emotion Recognition in the context of a job interview. Suppose that during an interview, an applicant's facial expressions are categorized as expressions of specific emotions, with special attention to anger, because it is assumed that people who feel and express anger at their place of employment are likely to be worse at collaborating with their fellow employees. After the interview, the interviewer reviews the report of the AI Emotion Recognition system on the applicant. The report claims that the applicant felt anger during the interview, and the interviewer rejects the applicant on this basis. Specifically, the interviewer reasons that the applicant may not be conducive to a collaborative workplace because of their tendency for anger.

This fictitious case represents actual use of the technology. It is increasingly common for employers to use AI in their hiring process (Wells, 2024). Job applicants are often required to participate in video interviews that undergo automated evaluation from the AI-powered software. Most notably, the company HireVue offered software to the effect described in the example: their system generated an employability score for each applicant based on evaluations of their facial movements, word choice, and speaking voice (Harwell, 2019).[5]

Assume that the AI Emotion Recognition system in question is accurate: the applicant really did feel anger during the interview. However, this fact alone is not a good reason to reject the applicant: feeling anger during an interview is not by itself evidence that the applicant is less suitable for employment than other applicants. Instead, the interviewer must also infer that the applicant's anger is inapt and that this is relevant to their job performance. If this inference is warranted, perhaps it would justify the interviewer's decision to reject the applicant in favor of other applicants: people who become angry in circumstances that do not call for such emotional reactions are not conducive to a collaborative workplace.

But suppose that the applicant was angry at the interviewer because the interviewer made a passing remark that disparaged the applicant on the basis of their social identity, such as their gender, race, or sexual orientation. In these circumstances, the applicant's anger is *apt*: it is a correct response to the genuine offensiveness of the interviewer's actions during the interview.

[5] HireVue claims that they no longer evaluate applicants based on facial analysis because this data did not offer much more insight into the applicant's employability than analysis of their speech (Knight, 2021). Although we think this is good news, it does not diminish the importance of the present project for three reasons. First, absent legal regulation and oversight, companies like HireVue may use the technology in the future for job interviews. Second, HireVue and other companies still offer assessments of emotion-related skills of applicants based on analysis of their speech, which may raise our concern about affective injustice in a new form. Finally, our concern about affective injustice may apply to other uses of AI Emotion Recognition (via facial analysis) beyond job interviews. For example, 4LittleTrees is a software suite that some schools in Hong Kong use to detect student emotions during learning and assessments (Chan, 2021). Ascertaining the aptness of student emotions may be required to treat students justly (and for desirable learning outcomes).

An emotion is apt only if (1) it is fitting or an accurate evaluation of its object, and (2) it is based on evidence that justifies that emotion.[6] To elaborate on the first necessary condition, in the same way that the accuracy of perception requires certain facts about the external world to obtain, the accuracy of emotions requires certain evaluative facts to obtain. Roughly, this is because the intentional objects of perception (such as redness) and intentional objects of emotion (such as offensiveness) are constituents of the relevant facts (Kenny, 1963; Deonna and Teroni, 2012).[7] So, an episode of anger is apt only if a corresponding evaluative fact obtains: that the object of that episode is offensive. To elaborate on the second necessary condition, the emotion in question must be felt in light of evidential reasons sufficient to justify that emotion.[8] For example, the applicant has perceptual evidence of the interviewer's microaggression, and the applicant's anger is based on this justificatory reason. But if the applicant had reason to doubt the reliability of their perceptual evidence, this would make their anger unjustified, and thus, inapt.

This analysis of apt emotions helps us understand what is at issue when the interviewer uses the report provided by the emotion recognition technology to conclude that the applicant is unsuitable. Whether the applicant's anger is evidence relevant to their suitability as an employee depends on whether their anger is in fact apt. This is because the feeling and expression of apt anger does not immediately provide a reason to believe that an applicant would be a difficult co-worker or to reject their application on this basis. The applicant who experiences apt anger

[6] Our project does not depend on spelling out the necessary and jointly sufficient conditions for apt emotions, and we do not attempt to do so here.

[7] Note that we do not claim here that emotions *represent* evaluative properties in the way that perceptual experiences represent color properties. Although *perceptualism* is a popular view of emotions (e.g., Tappolet, 2016), it has become increasingly disputed in the literature on emotions (e.g., Deonna and Teroni, 2015). In any case, our arguments here do not depend on settling this debate.

[8] Some argue that there are no reasons for emotions (e.g., Maguire, 2018). This threatens the analysis of apt emotions as fitting emotions that are supported by evidential reasons. Our project can be modified to accommodate this view. In brief, there are still reasons to regulate or manage unfitting emotions, even if there are not reasons for the emotions themselves, and judging that someone's fitting emotion is poorly managed is unjust (in the way described in the next section).

correctly registers the offensiveness of the interviewer's remarks. In other words, the fact that the interviewer made offensive remarks permits the applicant to be angry at those remarks, in the same way that a fearsome animal permits us to be afraid or an irrevocable loss of a loved one permits us grief. Of course, it is possible for someone to express apt anger in a morally objectionable way, such as verbally abusing or humiliating the object of one's anger. But the applicant expresses their apt anger through subtle changes in their facial expression, and it is not clear why expressing anger in this way is morally objectionable or otherwise evidence that the applicant is unsuitable for the position. In the next section, we characterize the type of mistake that the interviewer makes as a kind of injustice, done in this case to the applicant.

3.2. AI Emotion Recognition and Affective Indiscrimination

We argue that using AI Emotion Recognition in this way is not merely a failing wherein one judges beyond one's evidence in assuming whether someone's emotions are apt. It is also an instance of affective injustice or a type of wrongdoing with respect to our capacity for affective states including emotions. We adopt Pismenny et al.'s (2024: 154) definition of affective injustice: affective injustice occurs when "the treatment of emotions is unjust, or emotions are used to treat people unjustly," wherein injustice is understood as an arbitrarily imposed disadvantage. A *disadvantage* is a deprivation, including a demotion in status or social position, which is *arbitrary* to the extent that it is unjustified by the moral reasons relevant to the situation. For example, being denied a promotion on the basis of one's gender is an arbitrary disadvantage: a promotion is a valued resource with respect to its benefits to one's status and earnings, and one's gender is not a moral reason for the denial.[9]

[9] An alternative definition of affective injustice comes from Gallegos (2021). According to this understanding, a state of affective injustice is one in which individuals or groups are deprived of affective goods that they are owed. Gallegos identifies two fundamental affective goods: subjective (emotional) well-being and apt emotions. Although

To take a paradigmatic example of the phenomenon, consider Srinivasan's (2018) argument that victims of oppression have a right to their apt anger against their oppressors. Some argue that anger is counterproductive to ameliorating injustice, and so there is often a sufficient reason to regulate or manage that anger, despite its aptness. For example, if one acts in anger against one's oppressors, this may obscure the legitimacy of one's grievances to one's audience, because such audiences often assume that minorities that act in anger do so unreasonably (Pettigrove, 2012; Nussbaum, 2016). In response, Srinivasan argues this itself is affective injustice: in a just world, one would not need to make a tradeoff between one's apt feelings and one's interests (including justice for one's oppression) (2018: 131-136). Thus, considerations of affective justice give us a (prima facie) moral reason against suppressing apt anger or its expression. This affective injustice may be further characterized as an attempt to regulate someone else's emotions in ways that are ultimately harmful to that person or unproductive with respect to that person's interests (Archer and Mills, 2019).[10]

Here we focus on a specific type of affective injustice that we call *affective indiscrimination.* Someone commits affective indiscrimination when they form an incorrect doxastic attitude about whether an emotion is apt, when the emotion in question is in fact apt. In these circumstances, the correct doxastic attitude is a judgment that affirms the aptness of that emotion. Thus, it is incorrect to falsely believe that the emotion is inapt.[11] This is a type of

we speculate that the type of affective injustice of interest to us (affective indiscrimination) threatens apt emotions, we do not spell out our concerns here in terms of fundamental affective goods.

[10] Opponents of anger may attempt to reframe their critique to concede that it is morally permissible to feel but morally impermissible to express or act upon. Srinivasan (2018: 136-139) responds with an argument against the strict separation of emotional experience and emotional expression. We discuss related issues in section 4.2.

[11] Our account does not entail that an absence of doxastic attitudes may be affective indiscrimination, but what about suspending judgment (plausibly a fundamental doxastic attitude in its own right) about the aptness of an emotion? This turns on issues about the nature of suspension of judgment. In brief: if suspension of judgment is constitutive of inquiry (Friedman, 2019), then suspending judgment (about whether someone's emotion is apt) is not an incorrect doxastic attitude, as one has not settled the question about whether the emotion is apt. But if suspension

respect for an exercised capability: someone who feels an apt emotion is responding successfully to the evaluative facts.[12] As a result, victims of affective indiscrimination are denied the respect that they are due in responding to the evaluative facts in the correct way. So, like other types of affective injustice, affective indiscrimination is the mistreatment of someone's emotions through an arbitrary disadvantage to that person, who is deprived of due respect in the absence of any justifying moral reason given the aptness of their emotion.

For example, suppose that you are angry at a comedian's joke. If I judge that your anger is inapt, but in fact your anger is apt because the joke is offensive, then I have failed to discriminate between apt and inapt anger. In doing so, I have disrespected you as an affective being capable of responding to the evaluative world successfully, and more specifically, as someone who has successfully responded to the offensiveness of the joke in virtue of your anger. Similarly, if I treat you as if your anger is inapt, such as by rejecting your testimony about why the joke is offensive because I have judged that your anger is inapt, that treatment is also affective indiscrimination, and gives rise to the same type of disrespect.

Returning to the issue under discussion, we argue that using AI Emotion Recognition to categorize someone's emotions may facilitate affective indiscrimination insofar as it motivates doxastic attitudes that deny the respect due to apt emotions. This is an injustice because our emotions — including whether our emotions are apt — matter with respect to how others ought

---

of judgment is sometimes the end of inquiry, then suspending judgment (about whether someone's emotion is apt) is an incorrect doxastic attitude when it closes inquiry, as one has settled on a denial of the respect described here.

[12] What type of respect are we owed for apt emotions? We speculate that at least some (but perhaps not all) apt emotions demand appraisal respect, a type of respect that we are owed in virtue of some aspect of our moral character. For example, someone's apt anger may reflect well on their moral concern for themselves or others. See Darwall (1977) for more on appraisal respect. More generally, we can conceive of apt emotions as demanding evaluative respect, which is just to say that aptness is a standard by which we judge emotions. See Hudson (1980) for more on evaluative respect.

to regard us and act toward us.[13] For example, if the applicant's anger is apt, then they should be regarded as appropriately sensitive to the offense against them, and not as someone whose anger is out of touch, so to speak, with the evaluative facts about what is offensive. In this way, apt anger merits respect from others: it ought to be appraised as a correct response to an offense. For the interviewer to regard the applicant's anger as inapt and to reject the applicant on that basis is an injustice because the interviewer does not manifest the respect that the applicant deserves in light of their apt anger. Likewise for apt emotions generally. On the other hand, when someone's emotions are inapt, they should not be regarded as appropriately sensitive to the relevant evaluative facts, and so they should not be afforded such respect.

Before we continue to explore the relationship between affective indiscrimination and AI Emotion Recognition, let us pause to make three clarifications. First, although we conceive of affective indiscrimination as a distinct type of affective injustice, it is closely related to other types of affective injustice. In Srinivasan's (2018) view of affective injustice, apt anger may frustrate the interests of the oppressed in justice for oppression because the audience to the anger of the oppressed may assume that that anger is unreasonable in some way. In our view, that audience commits affective indiscrimination if they fail to distinguish apt from inapt anger, and in so doing, that audience perpetuates the tradeoff that the oppressed must make between their apt anger and their interests in justice. Consider also what Whitney (2018) calls *affective marginalization* wherein someone's emotions are not given the influence that is often afforded to others. It is not implausible that this marginalization may occur because of affective indiscrimination. For example, if a man does not accept that a woman's anger is apt, because he

[13] Some might wonder whether this kind of harm rises to the level of injustice. We believe that it does. As described above, affective injustice occurs when "the treatment of emotions is unjust, or emotions are used to treat people unjustly" (Pismenny et al., 2024: 154). When an individual's emotions are incorrectly appraised, it is not merely that they are harmed; they are arbitrarily disadvantaged on the basis of an incorrect emotional appraisal.

falsely judges that a woman's angry outburst is just a product of her bad mood and not about him in particular, then he has failed to respect the aptness of her anger. In light of these connections, we hope that isolating affective indiscrimination from other types of affective injustice enables us to understand the underlying moral issues more clearly. But we leave spelling this out in detail for a future project.

Second, we conceive of affective indiscrimination as affective injustice (and thus wrongful) regardless of its downstream consequences. Indeed, given the respect that we deserve for our apt emotional responses to the world, affective indiscrimination merely consists of judgments or other doxastic attitudes that fail to acknowledge apt emotions. In this way, affective indiscrimination is similar to *doxastic wrongdoing*, in which the wrongdoing against someone consists merely of our inappropriate doxastic attitudes toward them, such as believing that someone is less intelligent because of their race or gender (Basu, 2019). Nonetheless, it is important to recognize the harmful effects of affective indiscrimination. Recall the case of the applicant, where the interviewer's affective indiscrimination consists in regarding the job's candidate's apt anger as inapt. In addition to this affective injustice, further harm may be done when the applicant does not get the job and is thereby denied an opportunity for social and economic advancement. Consider the systemic nature of this affective indiscrimination and the resulting harm: if it is standard practice to use emotion recognition technology during interviews, then the applicant may risk losing out on a wide range of beneficial opportunities because of affective discrimination. If the applicant becomes aware of this risk, they face the same type of tradeoff that Srinivasan describes in her work: either feel and express their apt anger or manage that feeling and its expression so that it does not threaten their other interests. (This illustrates again our earlier point that affective indiscrimination can facilitate further affective injustice, in

addition to potential material harms.) Moreover, Archer and Mills have argued (2019) that there are significant harms or losses to managing emotions (especially anger). If this is correct, then when the applicant extinguishes their anger or suppresses its expression through emotional regulation techniques, they do not merely trade in their right for apt anger in exchange for the promotion of their social and economic interests. Rather, they must perform additional emotional labor to excise this anger, and likely at a cost to their psychological and physiological health. In this way, affective indiscrimination may cause a wide variety of harms.

Finally, note that affective indiscrimination does not necessarily result from a failure to believe according to one's evidence. In the previous section, we emphasized that the interviewer infers that the applicant is less suitable for the job than other candidates on the basis of the candidate's anger. The mere fact that the applicant is angry is not itself evidence that the applicant is less suitable for the job; he must also assume without any evidence that the applicant's anger is inapt. In this particular case, the interviewer commits affective indiscrimination because he does not believe according to his evidence. If he had, he may have continued inquiry about the aptness of the applicant's anger. Nonetheless, someone may believe according to their evidence but still commit affective indiscrimination. For example, I may have good reason to believe that your anger is inapt because all of my available evidence points in this direction, even though your anger is in fact apt. It is plausible here that my wrongdoing is unintentional in a way that may mitigate my moral responsibility and blameworthiness for it, but nonetheless I wrong you by denying you respect due to you in virtue of your apt anger.

4. Objections and Replies

4.1. Is AI Emotion Recognition Worse than the Status Quo?

One objection to this challenge to AI Emotion Recognition from affective injustice is that we are no better than AI in this respect. For instance, suppose that the interviewer noticed the applicant's anger himself, judged that the applicant's anger is inapt, and then decided to reject the applicant on this basis — all without input from AI Emotion Recognition. Indeed, the objector may agree that affective injustice is a serious problem in the context of a job interview but argue that our formulation of this critique incorrectly blames the technology for our (unjust) social norms about when we are permitted to feel and express emotions like anger.

In response, we maintain that the present challenge to AI Emotion Recognition is of special significance, though we agree that it is a continuation of existing patterns of affective injustice. To see why, let us begin with a comparison case. Consider how search engines like Google can perpetuate oppression against race and gender minorities with algorithms that reinforce harmful stereotypes about those minorities. For example, prior to 2012, using search terms like "black girls" on Google would overwhelmingly return pornographic images and websites, thereby reinforcing sexist and racist ideas (Noble, 2018). Although the stereotypes reflected here about black women existed long before search engines, its presence in new technology elicits at least two new ethical concerns. First, to the extent that we rely on search engines and consider them to be generally reliable, they take on a veneer of objectivity, and we may uncritically accept the information that it presents to us, including how to implicitly categorize (stereotype) others (and ourselves). Second, whereas it is relatively straightforward to understand how individuals can be morally responsible for sexist or racist actions, and it is even sometimes possible to pursue legal actions against them, it is not immediately clear who is morally responsible for a biased search engine algorithm, and there are no established legal mechanisms to hold anyone legally responsible or enforce a demand for change.

AI Emotion Recognition raises parallel concerns with respect to affective injustice. First, many users may come to regard AI Emotion Recognition as more objective than human attempts to either detect emotions or make decisions on the basis of those emotions. It is easy to see the appeal of trusting the technology — it offers the illusion of meaningful quantification in contrast to the idiosyncratic decision-making of human beings. For example, in the case of the job interview, it may generate a convenient 'employability score' for each applicant based in part on their emotional responses. Along these lines, a company that uses AI Emotion Recognition for its interviews may believe that the technology helps them make more objective decisions than relying on the judgments of an interviewer, and its apparent objectivity will seem vindicated if its use becomes pervasive. In this way, AI Emotion Recognition in its current form threatens not only to continue existing patterns of affective injustice, but to entrench them more deeply, making the injustice more difficult to notice as it becomes encoded in a technology that we may come to trust uncritically.

AI Emotion Recognition may seem trustworthy because it exploits our susceptibility to be cognitively seduced by quantification, which tends to cause feelings of clarity that are associated with epistemic achievements such as knowledge or understanding (Nguyen, 2021).[14] For example, the quantification of applicant employability into a numeral score that purports to represent the overall suitability of prospective employees may give rise to feelings of clarity about which applicant is most qualified for a position. This is presumably because quantified systems provide a high degree of *cognitive facility* or the ability to perform activities such as

[14] Nguyen (2021) investigates the *seduction of clarity* or how feelings of clarity may mislead us into thinking that we have considered enough evidence to cease inquiry into some question. Seduction of clarity via quantification is just one variety, albeit an important one to acknowledge in an epistemic environment of increasing quantification. See also Porter (1996) and Merry (2016) for related concerns about knowledge via quantification and its broad impact in domains such as science and politics.

communicating concise information, making decisions, or generating justifications (Nguyen, 2021: 245-246). Of course, the problem is that quantification may cause us to have feelings of clarity absent genuine knowledge or understanding, such as if the quantification is not a reliable indicator of the quantified phenomenon.

Second, AI Emotion Recognition complicates the issue of moral and legal responsibility. When someone does not respond correctly to your apt anger, such as by offering an irrelevant excuse on behalf of your offender, they are morally responsible for failing to respect your apt emotions (barring exempting circumstances). But when the interviewer carries out affective injustice by rejecting the applicant merely on the recommendation of his company's AI Emotion Recognition system, where does the moral responsibility for that affective injustice lie? With the interviewer? With the interviewer's company at large or some part of its management team? With the designers of that AI Emotion Recognition system? Some combination of the previous options? Unless the parties in question or legal authorities agree on the answer to this question about moral responsibility, it will be difficult to determine moral or legal responsibility for contentious uses of AI Emotion Recognition, including those that facilitate affective injustice. This is likely to stifle the development of public policy to regulate some uses of the technology, even when demonstrable harm or unfairness occurs as a result of those uses.[15]

4.2. Does Aptness Matter?

An objector may claim that someone's emotions are usually evidence relevant to making decisions about how to treat that person, regardless of whether those emotions are apt. For example, someone's anger may be a good reason to avoid them, even if that anger is apt, because

[15] See Matthias (2004) for an influential formulation of the problem of responsibility gaps that arises from autonomous systems. See also Da Silva (2024) for a recent overview of the literature on this issue, including suggestions about how to handle such gaps.

that person is hostile to everyone around them when angry, not just the object of their anger. In such cases, it is morally permissible to disadvantage people in some way in light of their emotions, regardless of their aptness. In the interview example, even if the applicant's anger is apt, it still violates norms of 'professionality' for the applicant to express that anger during the interview.[16] Such norms govern how one ought to act in the workplace as an employee. The objector may continue that the interests of affective justice do not permit us to express our emotions whenever we want and to whoever we want. Thus, in the case of the job interview, the applicant should express their anger in some other context other than during the interview itself to the interviewer.

Our response here is twofold. First, someone's emotions are predictive of their actions only in a limited way. Although emotions are typically accompanied by specific desires, how one chooses to satisfy that desire is open-ended. For instance, if anger is typically accompanied by a desire to intervene in some way against the putative wrongdoer, such as to punish the wrongdoer or make them aware of the seriousness of their wrongdoing, this desire may be satisfied in a wide variety of ways.[17] For example, the angry applicant may satisfy their desire to intervene against the interviewer by registering a formal complaint with the company about how they were treated during the interview. Thus, absent specific information about how the applicant expresses their anger, one should not believe that they will be a worse employee merely because of their anger. In fact, they may use that anger productively in defense of their company, product, or team.

[16] See D'Arms & Jacobson (2000) for an influential argument that the fittingness and appropriateness of emotions are distinct. They persuasively claim that an emotion may be fit without being appropriate. Note here though that our present discussion concerns the norms of emotional expression, and not whether the emotion itself is appropriate.

[17] What is the relationship between anger and motivation for specific types of actions? The orthodox view is that anger motivates us to take revenge or retribution (Pettigrove, 2012; Nussbaum, 2016). However, others reject this characterization and instead propose that anger (at least sometimes) motivates us to seek recognition of wrongdoing from the wrongdoer, not necessarily payback (Srinivasan, 2018; Silva, 2022). We do not take a stand here on this debate but instead claim that anger leaves us with significant latitude in action regardless of how it motivates.

Second, we agree that affective justice does not permit us to express our apt emotions however we wish. There are often moral reasons against expressing apt emotions in particular ways or at particular times. Indeed, it would be unprofessional (not to mention morally wrong) for the applicant to threaten the interviewer because of the interviewer's microaggression. But the applicant only expresses their anger through their facial expression, and it is not obvious that it is always unprofessional to express anger in the workplace. For instance, a team building exercise may involve team members expressing their anger to each other to learn to respond empathetically in order to work better together. In addition, if the interviewer has already violated norms of professionality with a microaggression, the applicant may be freed of any obligation to continue to be professional in that setting.

The objector may point out that specific professions require successful employees to suppress expression of apt emotions. For example, a flight attendant or airline steward is expected to maintain an overall attitude of hospitality toward passengers in ordinary circumstances, even if it is apt to feel frustrated or angry at them. Similar expectations apply for jobs in the service industry more generally.[18] If the demands of a specific profession require suppressing the expression of apt emotions, and AI Emotion Recognition can help assess whether an applicant suppresses successfully, then rejecting an applicant is not necessarily motivated by affective indiscrimination or the disrespectful dismissal of their apt emotions. Moreover, if AI Emotion Recognition is predominantly used to ascertain whether people can suppress their apt emotions, such as evaluating a job interview for a position in the service industry, then our claim that the technology facilitates affective injustice seems misguided.

[18] There are additional moral issues here related to requiring this type of emotional labor in the workplace, especially from people who work in the service industry, that are outside of the scope of this paper. See Hochschild (2012) for one influential discussion of this. We are open to the view that affective injustice may occur in the workplace more generally than we indicate here, especially in virtue of expectations to suppress apt emotions.

We maintain that such use of AI Emotion Recognition would risk affective injustice. It is reasonable to expect civility and hospitality from employees, especially in service professions, and even in the face of offensive co-workers or customers, but a successful employee is not (and should not be) required to suppress anger entirely, which would entail (among other implausible feats) perfectly managing one's facial expressions. Suppression to this degree is beyond normal human capability, and the pressure to suppress apt anger entirely expresses disrespect for its aptness. So, in response to the objector, we claim that professional standards do not provide moral justification for treating the apt anger of employees without respect (by pressuring them to suppress their anger entirely).

Moreover, it would be undesirable for employers to use AI Emotion Recognition to ascertain whether employees can suppress their apt emotional expressions. Arguably, the ideal employee expresses their apt emotions in a manner conducive to their professional role, and jobs in the service professions are no exception. For example, a flight attendant may express their apt anger to a passenger whose conduct is endangering others through assertive language, a stern tone of voice, and a steely expression. Conversely, a flight attendant who does not effectively express anger in these circumstances may fail to elicit compliance from passengers. So, the objection overstates the value of AI Emotion Recognition in screening for employees that can suppress their apt emotions.

4.3. Is Affective Indiscrimination Truly Wrong?

We argue that affective indiscrimination constitutes affective injustice or wrongdoing: to falsely believe that another person's apt emotions are inapt wrongs that person. Assuming that we often have false beliefs about the aptness of emotions, it follows from our view that we often commit (or are victims of) this type of wrongdoing. This prompts another type of objection: our

account entails that there is an implausibly high amount of wrongdoing.[19] Here we will detail and respond to a series of three objections along these lines.

On the first objection, we should reject moral views that entail that there is an excessive amount or frequency of injustice because this threatens the meaningfulness of the distinction between justice and injustice. If every case of affective indiscrimination is a case of affective injustice, and we often have false beliefs about whether emotions are apt, then injustice arises so often that we should doubt any moral view that entails this.

However, this seems like an arbitrary restriction on the types of injustice that may exist. Moreover, the frequency of a type of injustice is a contingent matter. For example, imagine a society where theft is much more common. This by itself would not make theft less of an injustice. Unless a moral view entails that *every* action is just or unjust, the meaningfulness of the distinction is not obviously at stake. Of course, it may be helpful to evaluate the severity of unjust actions, and this may speak to an underlying concern that excessive injustice makes it difficult to deliberate about what to do in response. The relative severity of injustices may guide our deliberation about which injustices to prioritize addressing with limited resources, but our present project is to illuminate an undertheorized injustice, not to make claims about its relative severity or urgency as an injustice.

On the second objection, our view is inconsistent with the commonplace observation that living well with other people requires the frequent suppression of apt emotional expressions. For example, it may be rude for us to express our disgust at your coffee breath, even though our disgust is apt, and you may take actions that disadvantage us because of our expression of apt

[19] We thank an anonymous reviewer for raising this objection.

disgust. This example demonstrates that for one to be merely disadvantaged because of one's apt emotions is insufficient to establish that one has suffered an injustice.

We maintain that this objection ultimately misunderstands affective injustice and our view about affective indiscrimination. Recall that affective injustice occurs when a disadvantage is imposed on someone because of their emotions that is arbitrary and so unjustified by moral reasons. But in the example, there are putative moral reasons to disadvantage someone who expresses apt disgust, which are perhaps grounded in our social norms of etiquette that help us display sensitivity in choosing which emotions to express and how. More specifically, our proposal is that affective indiscrimination is affective injustice: you wrong someone whenever you falsely believe that their apt emotion is inapt. But this is not at issue in the example. You need not believe that our disgust is inapt in order to respond with hostility. In other words, you do not commit affective indiscrimination against us when you respond negatively to our expression of apt disgust.

On the third objection, it is our reactions (or lack thereof) to affective indiscrimination that tell against our view. If affective indiscrimination constituted genuine wrongdoing, then given its frequency, we would blame each other for it frequently (or at least recognize that blame would be a correct reaction). But we do not in fact blame each other frequently for affective indiscrimination (or recognize that blame would be a correct reaction). For example, when someone falsely believes that their partner's anger against them is inapt, their partner may seek to correct that mistake without necessarily blaming them for affective indiscrimination.

However, this only demonstrates that some affective indiscrimination is blameless, not that it is morally permissible. As noted earlier, it is plausible that affective indiscrimination is blameless when one's belief is supported by one's overall evidence. Along these lines, consider a

different version of the example. Suppose that someone falsely believes that their partner's anger is inapt, despite evidence that they wronged their partner by breaking a promise to them. In these circumstances, they are blameworthy for affective indiscrimination against their partner. Not only have they failed to respect their partner's apt anger by responding with an inappropriate doxastic attitude, but they do so despite having epistemic reasons to adopt the appropriate doxastic attitude.

4.4. Could Technological Improvements Ameliorate the Problem?

In our discussion about the accuracy challenges to AI Emotion Recognition, we briefly suggested how improvements to the technology may improve its accuracy and thereby overcome objections to its use. Could the same be said about the present challenge of affective injustice? Is it possible to develop AI Emotion Recognition that identifies not only which emotion someone is feeling, but also determines whether that emotion is apt?

We are not nearly as optimistic about the future capabilities of AI Emotion Recognition here as we were in the case of the accuracy challenges. Recall that an emotion is apt only if the object of that emotion has the evaluative property that corresponds to that emotion. So, in order to assess whether an emotion is apt, AI Emotion Recognition would require data about the object of the emotion. But it is very difficult to see what data would allow AI Emotion Recognition to predict with accuracy the object of emotions, short of testimony from the person whose emotion it is. For example, suppose that in the course of the interview, the applicant is reminded of a difficult experience at a previous job, and they feel apt anger at being mistreated in a context outside of the interview altogether. If AI Emotion Recognition only collects data from the interview itself, and during the interview there is no information indicating that the applicant's

anger is about something from another context altogether, then assessing the aptness of emotions seems to run up against principled limitations of the technology.[20]

Nonetheless, even if it turns out that AI Emotion Recognition can assess whether an emotion is apt, it does not impugn the significance of the present challenge to AI Emotion Recognition from affective injustice. Until these hypothetical developments are made, the current state of the technology does not provide us with evidence about the aptness of emotions that it detects, and so there is still a reason to examine our present use of AI Emotion Recognition to ensure it does not facilitate affective injustice.

5. Concluding Remarks: How Can We Use and Design AI Emotion Recognition to Avoid Affective Injustice?

We have argued here that AI emotion recognition facilitates affective indiscrimination, especially when someone's emotions are judged to be inapt when they are in fact apt. We claim that this affective indiscrimination is not only a case of affective injustice, but also that it may cause further harm and unfairness.

We hypothesize that AI Emotion Recognition facilitates affective indiscrimination through its design and marketing. Because this technology provides us with measurable data, it encourages us to transform difficult questions into simple procedures. For example, the question of whether to hire an applicant transforms into a procedure to hire the candidate with the highest employability score (as measured in part by the emotions expressed by that applicant during their interview). In this way, the general approach here by designers, advertisers, and users reflects an

[20] Perhaps a hypothetical technology could scan someone's brain to assess the objects of one's mental states, but it is highly contentious whether this type of information about someone's brain could be used to infer what their mental states are about. Moreover, the resulting technology would be significantly different from AI Emotion Recognition in its current form, which uses information downstream from emotions to identify them (such as facial expressions thought to be caused by the emotion in question).

attitude of *technological solutionism*, or the idea that solutions to our social problems can rely exclusively or heavily on technological improvements rather than through revisions to our social institutions and practices (Morozov, 2013).

That being said, we have not claimed that emotion recognition technology is inherently unjust.[21] Instead, we conceive of our challenge as an invitation to approach our understanding of this technology and its limitations in a new light. Along these lines, we see significant value in merely recognizing emotions that may otherwise be overlooked. For example, suppose that a therapist is using emotion recognition to help identify the emotions of a patient during a session. In this case, the therapist may withhold making any judgments about whether those emotions are apt, but nonetheless encourage their patient to recognize those emotions and then to reflect on them. Similarly, an educator may use emotion recognition to notice a student's frustration during learning, and then discuss that frustration with the student, rather than make interventions to their learning that presuppose the aptness of that frustration.[22]

To generalize our approach from these examples, we suggest that there is a moral reason to use emotion recognition technology in such a way as to help people recognize their own emotions. Once someone recognizes their own emotions, then they can help decide what type of intervention is required (if any). For instance, the teacher in the previous example may work with

[21] We also do not claim that we have exhausted the moral issues that it raises. Another important ethical consideration against widespread or public use of emotion recognition AI concerns our privacy (McStay, 2020).

[22] Vold and Hernández-Orallo (2021) discuss the possibility of neurodiverse people using AI Emotion Recognition technology (e.g., smart glasses) to detect and respond to neurotypical social cues. There are certainly benefits to this kind of application: it could help improve communication between neurodiverse and neurotypical individuals, and it could even promote the autonomy of neurodiverse individuals, as they could rely less on therapists and support staff in social settings. However, a worry is that such technology would pressure neurodiverse individuals to conform to neurotypical social preferences, instead of encouraging a more mutual understanding. The issues here plausibly involve affective injustice: the emotions of neurodiverse individuals are often misinterpreted by neurotypical individuals, and neurodiverse individuals are exploited when they alone are pressured to do emotional labor to understand neurotypical individuals (as opposed to mutual understanding).We thank an anonymous reviewer for pointing out this potential application of the technology.

the frustrated student to identify the role that frustration is playing in the learning process and whether the student needs help managing that feeling. These decisions may be based on whether those emotions are apt, rather than settled by the mere fact that they feel that emotion. To make such decisions on the basis of the fact that someone feels an emotion without their input increases the risk of affective indiscrimination. This is because the person whose emotion it is may have relevant information about the aptness conditions of the emotion and whether it is apt.

To summarize, we propose that humans need to be kept in the loop, so to speak, when it comes to emotion recognition technology. We can use this technology to better understand other people, and to help them understand themselves, but only if they are given input about what the technology reports, and that report is supplemented with someone's own perspective on their emotions. Developing emotion recognition technology in this way may facilitate its benefits in addition to avoiding affective injustice.[23]

[23] For helpful conversations about this paper, we would like to thank Robert Howell, Michael Lynch, Garrett Mindt, and Susan Schneider. The authors would also like to thank audiences at Rice University and Bonn University for their comments and questions. We are also indebted to the pair of anonymous reviewers of this journal for helping us to make significant improvements to the paper.

References

Affectiva.com. (2018). *Solutions*. https://www.affectiva.com/what/products/

Archer, A. & Mills, G. (2019). Anger, Affective Injustice and Emotion Regulation. *Philosophical Topics*, *47*, 75-94.

Bandes, S. A. (2014). Remorse, demeanor, and the consequences of misinterpretation. *Journal of Law, Religion and State, 3*, 170–199.

Baron-Cohen, S., Golan, O., Wheelwright, S., & Hill, J. J. (2004). Mind reading: The interactive guide to emotions. Jessica Kingsley.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*, 241–251.

Barrett L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, *20*, 1–68.

Basu, R. (2019). What we epistemically owe to each other. *Philosophical Studies 176*, 915–931.

Buolamwini, J. & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, *81*, 1-15.

Chan, M. (2021). This AI reads children's emotions as they learn. CNN Business. https://edition.cnn.com/2021/02/16/tech/emotion-recognition-ai-education-spc-intl-hnk

Crivelli C., Russell J. A., Jarillo S., Fernández-Dols J. M. (2017). Recognizing spontaneous facial expressions of emotion in a small scale society of Papua New Guinea. *Emotion*, *17*, 337–347.

D'Arms, J., & Jacobson, D. (2000). The moralistic fallacy: on the 'appropriateness' of emotions. *Philosophy and Phenomenological Research*, 65-90.

Darwall, S. L. (1977). Two kinds of respect. *Ethics*, *88*(1), 36-49.

Da Silva, M. (2024). Responsibility gaps. *Philosophy Compass*, 19(9-10).

Deonna, J., & Teroni, F. (2012). *The Emotions: A Philosophical Introduction*. Routledge.

Deonna, J., & Teroni, F. (2015). Emotions as Attitudes. *Dialectica, 69,* 293-311.

Duran J. I., Reisenzein R., Fernández-Dols J.-M. (2017). Coherence between emotions and facial

expressions: A research synthesis. In Fernández-Dols J.-M., Russell J. A. (Eds.), *The science of facial expression* (pp. 107–129). Oxford University Press.
Ekman P. (1980). *The face of man*. Garland Publishing, Inc.
Fernández-Dols J.-M. (2017). Natural facial expression: A view from psychological construction and pragmatics. In Fernández-Dols J.-M., Russell J. A. (Eds.), *The science of facial expression* (pp. 457–478). Oxford University Press.
Friedman, J. (2019). Inquiry and Belief. *Noûs, 53*, 296-315.
Gallegos, F. (2021). Affective injustice and fundamental affective goods. *Journal of Social Philosophy, 53*(2), 185-201.
Gunnery S. D., & Hall J. A. (2014). The Duchenne smile and persuasion. *Journal of Nonverbal Behavior, 38*, 181–194.
Harwell, D. (2019). *A face-scanning algorithm increasingly decides whether you deserve the job.* The Washington Post. https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/
Hill, K. (2022). *Microsoft Plans to Eliminate Face Analysis Tools in Push for 'Responsible A.I.'* The New York Times. https://www.nytimes.com/2022/06/21/technology/microsoft-facial-recognition.html
Hochschild, A. R. (2012). *The Managed Heart: Commercialization of Human Feeling*. University of California Press.
Hudson, S. D. (1980). The nature of respect. *Social Theory and Practice*, *6*(1), 69-90.
Johnson, K. (2023). *Real-time empathetic marketing is starting a retail revolution. Is emotion AI leading the way?* The CEO Magazine. https://www.theceomagazine.com/business/innovation-technology/emotion-artificial-intelligence/
Kaye, K. (2022). *Companies are using AI to monitor your mood during sales calls. Zoom may be next.* Protocol. https://www.protocol.com/enterprise/emotion-ai-sales-virtual-zoom
Kenny, A. (1963). *Action, Emotion and Will*. Routledge and Kegan Paul.
Kim, E., Bryant, D., Srikanth, D., Howard, A. (2021). Age Bias in Emotion Detection: An Analysis of Facial Emotion Recognition Performance on Young, Middle-Aged, and Older Adults. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society.* New York, USA.
Klare, B., Burge, M., Klontz, J., Vorder Bruegge, R., Jain, A. (2012). Face Recognition Performance: Role of Demographic Information. *IEEE Transactions on Information Forensics and Security*, *7*, 1789-1801.
Knight, W. (2021). Job screening service halts facial analysis of applicants. *WIRED*. https://www.wired.com/story/job-screening-service-halts-facial-analysis-applicants/
Kyriakou, K., Kleanthous, S., Otterbacher, J., Papadopoulos, G. (2020). Emotion-Based Stereotypes in Image Analysis Services. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) (UMAP '20 Adjunct). New York, 252-259.
Maguire, B. (2018). There are no reasons for affective attitudes. *Mind*, *127*(507), 779-805.
Martinez-Martin, N. (2019). What Are Important Ethical Implications of Using Facial Recognition Technology in Health Care? *AMA Journal of Ethics, 21*(2), 180-187.
Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and information technology*, 6(3), 175-183.

McStay, A. (2020). Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society*, *7*(1), 2053951720904386.
Merry, S. (2016). *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. University of Chicago Press.
Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. Public Affairs.
Nguyen, C. T. (2021). The seductions of clarity. *Royal Institute of Philosophy Supplements*, *89*, 227-255.
Noble, S. (2018). *Algorithms of Oppression*. NYU Press.
Nussbaum, M. C. (2016). *Anger and Forgiveness: Resentment, Generosity, Justice*. Oxford University Press.
Pettigrove, G. (2012). Meekness and 'Moral' Anger. *Ethics, 122,* 341-370.
Porter, T. (1996). *Trust in Numbers*. Princeton University Press.
Rhue, L. (2018). Racial Influence on Automated Perceptions of Emotions. *CJRN: Race & Ethnicity (Topic).*
Pismenny, A., Eickers, G., Prinz, J. (2024). Emotional Injustice. *Ergo, 11*(6): 150-176.
Pymnts.com. (2024). *AI Explained: Emotion AI Tries to Interpret Human Feelings.* https://www.pymnts.com/artificial-intelligence-2/2024/ai-explained-emotion-ai-tries-to-interpret-human-feelings/
Rychlowska M., Jack R. E., Garrod O. G. B., Schyns P. G., Martin J. D., Niedenthal P. M. (2017). Functional smiles: Tools for love, sympathy, and war. *Psychological Science, 28*, 1259–1270.
Silva, L. (2021). Anger and its desires. *European Journal of Philosophy*, *29*(4), 1115-1135.
Srinivasan, A. (2018). The Aptness of Anger. *Journal of Political Philosophy, 26*, 123-144.
Thomas, D, (2018). The cameras that know if you're happy - or a threat. BBC. https://www.bbc.com/news/business-44799239.
Tappolet, C. (2016). *Emotions, Values, and Agency*. Oxford University Press.
Vold, K. & Hernández-Orallo, J. (2021). AI Extenders and the Ethics of Mental Health. In Jotterand, F., Ienca, M. (Eds.), *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues* (pp. 177-202). Springer.
Wells, R. (2024). 65% of employers to use AI to reject candidates in 2025. *Forbes*. https://www.forbes.com/sites/rachelwells/2024/10/27/65-of-employers-to-use-ai-to-reject-candidates-in-2025/
Whitney, S. (2018). Affective Intentionality and Affective Injustice: Merleau-Ponty and Fanon on the Body Schema as a Theory of Affect. *The Southern Journal of Philosophy, 56*, 488-515.
Zetlin, M. (2018). *AI Is Now Analyzing Candidates' Facial Expressions During Video Job Interviews*. Inc.com. https://www.inc.com/minda-zetlin/ai-is-now-analyzing-candidates-facialexpressions-during-video-job-interviews.html
Zhao, L. & Tsai, R. (2011). *Locking and unlocking a mobile device using facial recognition.* Google Patents. https://patents.google.com/patent/US20120235790A1/en