Ethical considerations in emotion recognition technologies: a review of the literature

Amelia Katirai
Research Center on Ethical, Legal, and Social Issues, Osaka University
a.katirai.elsi@osaka-u.ac.jp

Abstract
As the global market for emotion recognition technologies (ERT)—which claim to use artificial intelligence to recognize emotions—rapidly expands, there is also increasing concern about their ethics. This paper reports on the results of a structured review of the literature on the ethics of emotion recognition technologies, to synthesize the ethical concerns expressed in the analyzed corpus of literature. This exploratory review draws on literature retrieved from the academic database Web of Science and from a hand-searching process, with a total of 43 articles included following a four-phased screening process. Three key areas of ethical concern were extracted from the literature: first, the risk of biased and unfair outcomes due to the faulty bases and problematic premises of ERT; second, the sensitivity of emotion data used by ERT; and third, the risk of harm that arises from the technologies in consequential settings including employment, education, healthcare, and policing. This paper additionally reports on a qualitative synthesis of the guidelines for ethical use of emotion recognition technologies proposed in the literature, finding that they address the need for both ethical design and implementation, and are focused most heavily on the need for: a defined scope for the use of ERT, ethical decision-making, fairness and non-discrimination, and privacy. Ultimately, this review finds that these technologies raise significant—and potentially insurmountable—ethical issues, even as their commercial development for widespread use continues.

Keywords
Emotion recognition technologies, ethics, artificial intelligence, literature review

Introduction
An increasingly lucrative market for emotion recognition technologies (ERT) continues to expand, with the implementation of the technologies proceeding apace. Estimates vary for the value of the global market for ERT, though there are indications that the industry may be worth as much as 37 billion USD by the year 2026 (1). ERT are part of a broader, multidisciplinary field often referred to as affective computing, encompassing "systems and devices that can recognize, interpret, process, and simulate emotion or other affective phenomena" (2,3). Though affective computing is a field made up of a range of technologies with diverse capabilities, the label is often used interchangeably—and at times misleadingly—with artificial emotion(al) intelligence, emotion(al) AI, affect recognition, and emotion recognition to refer to related technologies (3,4). The focus of this paper is on technologies that claim to use narrow artificial intelligence (AI) to identify emotions based on biometric signals, even when they are simply

detecting surface-level changes in facial movements (5). Therefore, "emotion recognition technologies" (ERT) is used as a standard term (6).

Whereas ERT were previously primarily designed for health-related applications, there has been a growing shift towards commercial uses, and the risk of a shift from recognizing, towards predicting, and ultimately potentially controlling behavior (2,7–9). ERT—like other technologies relying on biometric data (10)—is controversial, but this controversy is deepened by a lack of consensus around what emotion is, and robust critique of the proposition that it is possible for emotions to be "recognized" by machines (11,12). Thus, there is a small but growing body of literature expressing concerns about the ethics of emotion recognition technologies, even as major investments into their development and implementation continue. This paper adds to prior research (e.g., 13,14) by reporting on the results of a structured review and synthesis of this body of literature, which was conducted to scope the landscape around the ethics of ERT, and to identify the extent to which the ethical issues associated with ERT are prohibitive, in part through the application of two ethical frameworks in the discussion.

As a result of the analysis of the literature, this paper reports on three key areas of concern: first, the risk of biased and unfair outcomes due to the faulty bases and problematic premises which underpin ERT; second, the sensitivity of emotion data; and third, and the risk of harm arising from the use of ERT in consequential settings. Multiple studies included in this review proposed guidelines and principles for the ethical design and use of ERT. Therefore, this paper additionally reports on a qualitative synthesis of these guidelines. Moreover, the discussion of this paper extends beyond the literature identified in the reported review to draw in broader perspectives on the ethical issues raised by AI systems broadly, and to apply these to the case of ERT. Ultimately, though some argue for the possible benefits of responsibly designed and deployed ERT (5,15), the results of this review highlight persistent, significant, and potentially insurmountable ethical issues arising from ERT.

An overview of ERT

ERT are necessarily underpinned by certain assumptions about what emotion is, despite ongoing debate in this area. As indicated by Stark and Hoey (13), emotion may best be understood as a "compound phenomenon," made up not only of the physiological, expressive, and behavioral components read by ERT, but also by subjective and contextual components (3,8,13,16). It is noteworthy that, although at times used interchangeably, emotion and mood can be understood as subcategories of an overarching category of affect; emotion tends to have a shorter duration, and is more likely to have an object than mood (17).

As described by Bakir et al. (15) the origins of ERT can be traced back to the 19th century, when Duchenne de Boulogne "created a taxonomy of facial expressions that informs modern computer vision techniques." Today, they are being developed within these clusters both by relatively established technology companies such as Amazon, Hitachi, and NEC, and also by start-ups including Sensum, Affectiva, and Real Eyes (18,19). A bibliometric analysis conducted by Ho et al. (2) found two regional clusters of research activity in relation to affective computing more broadly: an Asia-Pacific cluster made up of the United States, China, Singapore, Japan, and India; and a European cluster, made up of Germany, the United Kingdom, and the Netherlands; Canada and Israel are also leaders in research and development (16).

It is noteworthy that ERT can be "bundled" with existing facial recognition systems, and are increasingly being applied in a variety of settings, as reported in Table 1 below:

Table 1 Areas of implementation of ERT

Area
Customer service (4,8)
Education (2,4,8,9,13–15,18,20)
Employment (2,4,13–15,21)
Entertainment (2,4,8,9)
Finance (8)
Governance and politics (8,20)
Healthcare (4,8,9,13,15,22)
Insurance (9)
Law enforcement and defense (4,8,13–15)
Marketing (4,8,9,14,15)
Mental health (4,13,14)
Social and other media (4,15,20)
Transportation (2,4,8,13–15)

The wide-ranging application of ERT as seen above, including in consequential settings such as education, employment, and law enforcement, brings to light an urgent need to better understand their ethics.

Purported benefits of ERT

A range of purported benefits arising from the use of ERT have been put forward—a few key exemplars are given here. It has been incorporated into systems in order to detect users' emotions and adjust the system itself to users' emotions (23). For example, ERT are proposed for use in advertising and retail in order to adapt the retail experience to the individual, providing customized recommendations to individuals based on their affective states, or offering augmented reality elements adapted to individual customers at a given point in time (24) . Though this level of customization may be perceived to be invasive, it is justified as providing "higher levels of service" to customers, "akin to 'living in a small town where everybody knows your name'" (24).

In other instances, the use of ERT has been proposed as a way to support health and well-being, including by assisting in the emotional self-awareness of individuals, often with the justification that greater awareness of emotional states will contribute to the development of a "better society" (24). Uses in healthcare are argued to be a way to increase personalization of healthcare and thus benefit patient health (25). A key use case for ERT in pediatric healthcare is said to be support for emotion recognition and emotional awareness in children with autism (23,26,27), and has been one of the earliest use cases proposed for affective computing broadly (28). Uses in this area include the incorporation of ERT into "serious games," through which children are prompted to identify an emotion, express it, and then identify it "in the wild" (23), or through the

development of a smart "emo-mirror" to aid in emotion detection (29). It is noteworthy, however, that an initial deployment study of the emo-mirror found concerns from health professional participants about the insufficiencies of such an approach, including the limitations inherent in a machine-based approach, the overly limited range of emotions such technologies are designed for, or the discomfort children may experience in using the mirror (29).

ERT is also put forward as beneficial to adults in the workplace. It is justified as a means through which employers can ensure the mental well-being of their employees (24), including "safeguard[ing] against toxic practices in the workplace" (30) by, for example, identifying indicators of harassment. Meanwhile, there are expectations that the introduction of AI tools to assist with the regulation of emotions could contribute to increases in productivity and effectiveness (31). A notable site for workplace surveillance of emotions using ERT is on the road, to improve safety as well as the mental health of drivers such as truck drivers, though the uses for ERT extend to private vehicles as well (32–34). Beyond this, it is also proposed as a way to manage crime and assist in policing, though there are concerns that it may simply be a modern day version of phrenology (1,20). In this way, then, the development and implementation of ERT across a range of settings are justified as bringing benefits to user experience, well-being, and safety. Yet, as will be explored below, the technologies themselves rest on fundamentally problematic assumptions.

Conceptualizing emotion

These uses for ERT fundamentally draw on an assumption that it is possible for emotional states to be detected from visual or biological cues, which has been strongly refuted (11). Though a range of models for conceptualizing emotion exist (18), including hybrid models, Stark and Hoey (13) identify three models of emotion in particular: emotion as "felt experience," through which emotions are primarily identified based on how they "feel" to the individual; emotion as an "evaluative signal," through which emotions cause or are caused by particular cognitive states; and emotion as a "motivating drive." ERT are based on this third model of emotion as a "motivating drive" (13). This model draws on the Basic Emotions approach proposed by Paul Ekman and colleagues, which claimed to identify six basic emotions as universal; therefore, emotions were understood to be primarily physiological and as something that could not be faked (35) . In that sense then, emotions are understood to be perceptible from the outside, making them potentially "machine-readable" (9,36).

Yet, this model—and the ERT developed based on it—has received robust critique over the faulty "science" (21) behind it, and for its overly limited understanding of emotion. It neglects the performative role of emotion (18), and the role of contextual and subjective factors, as will be discussed below (4,10,14,16). For this reason, this approach has been critiqued, notably by Barrett et al. (11), in a comprehensive review of the field which found that emotions indeed cannot reliably be identified based on facial movements. Moreover, the approach has been critiqued for drawing heavily on interpretations of emotions elicited in artificial settings, and thus being fundamentally unsound (11,37).

In light of the unfeasibility of reliably detecting emotions based on physical cues, it is noteworthy that the data used for emotion recognition is at best understood as proxy data (12), though there is a lack of consensus on associations between physiological cues and emotional

states (24). The data used for ERT includes facial expressions, physiology and vital signs, posture, gait, behavioral patterns, as well as text and speech, though much of the field remains based primarily on facial expressions, which is thus the focus of this paper (4,6,14,20).

Methodology

An exploratory, structured review of the literature was conducted to identify academic and grey literature articles on the ethics of ERT. A list of keywords related to ERT was compiled based on recent research in the area and were combined with terms related to ethics and ethical, legal, and social issues (ELSI) to create a search strategy. Web of Science was selected as the primary academic database due to its comprehensiveness. The search strategy was tested through a series of pilot searches to identify the optimal combination of keywords to maximize the retrieval of relevant articles. This search served as a base-point for gathering relevant literature, and was supplemented by a review of the references of all included articles to identify any additional relevant titles, and by hand-searches through Google Scholar and Semantic Scholar. The time period covered by this study was a 10-year period with a cut-off point of items published prior to June 15, 2022.

The retrieved articles were screened through a four-phased process: first, title screening of all retrieved articles; second, abstract screening for articles included in the first phase; third, full-text screening to determine final inclusion; and fourth, reference screening and hand-searching to identify additional articles of relevance (see Fig 1). Articles were included if there was a clear focus on the ethics of ERT. Articles focused on affective computing more broadly, such as those focused on AI or robots which simulate emotion, as well as articles focused on sentiment analysis and the recognition of emotion from text and other non-physiological sources were determined to be outside of the scope of this review and were excluded (e.g. (38)). Articles were excluded when they focused primarily on new technologies or techniques or otherwise addressed ethical issues only incidentally. Editorials and book reviews were also excluded.

Data about all included articles, including the author, aim, setting, methodology, findings, and sources of funding, were extracted into an expanded version of the table included in the Appendix. Key points from the articles were compiled thematically. The points were categorized through an inductive approach drawing on an adapted version of thematic analysis (e.g., (39)), identifying overarching themes emerging from the data. Continuing this inductive process, then, the thematic frame was progressively adjusted, until all themes were collapsed into three overarching themes, presented below. The aim of the analysis was to draw out the key themes across the literature to provide an overview of the landscape, rather than to categorize the literature or the issues into definitive categories. Therefore, as noted in the table of results included in the Appendix, there was overlap among the themes within particular articles, as multiple themes frequently appeared in a single article. When this occurred, single articles were referenced within more than one theme.

Multiple articles included in this study proposed a set of guidelines, postulates, or principles for the ethical implementation of ERT. As a secondary analysis, these guidelines were analyzed to identify commonalities across multiple guideline sets. The guidelines were extracted from their respective articles and coded by the author using a qualitative approach. This approach was

primarily based on inductive, open-coding inspired by thematic analysis (39). Coding was repeated over multiple rounds and with a period of time in between to ensure intra-coder reliability. Following inductive coding, reference was made to Fjeld et al.'s (40) analysis of guidelines for AI ethics in collapsing smaller themes into overarching themes (39). Following this, and in order to grasp the nature of the guidelines and which parts of the ERT lifecycle they were primarily relevant to, the themes were then clustered into two broader areas depending on whether they dealt with imperatives for design or for implementation.

Results

A total of 1,455 articles were retrieved through the search strategy described above. Through the four phases of screening, a total of 43 articles were included (Fig 1).

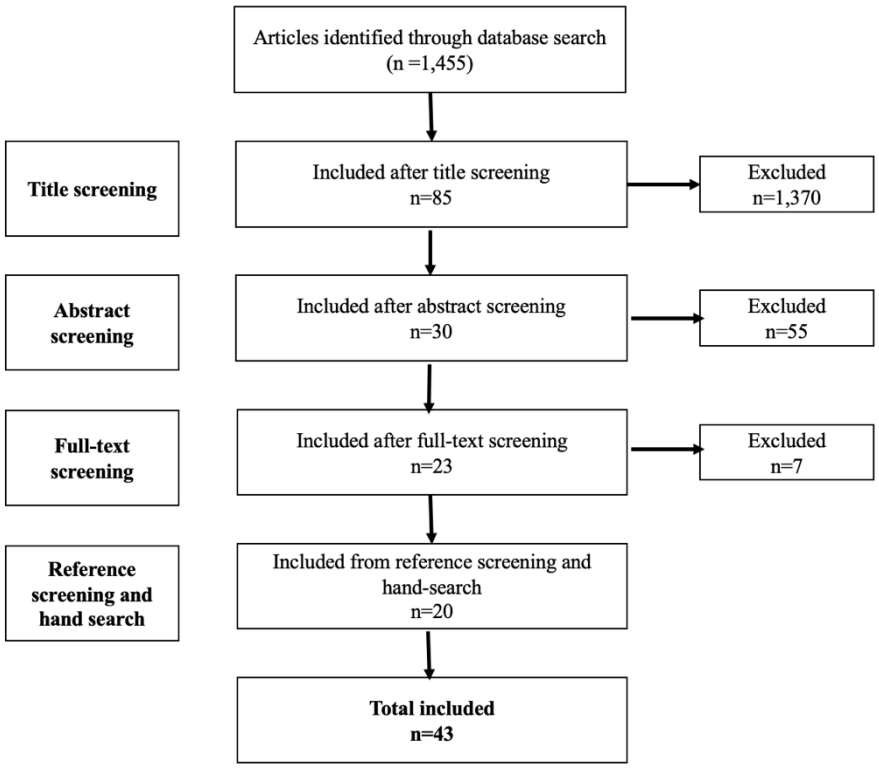

Fig 1 Flow of inclusion

The included works, their aims, and a representative extract or other key takeaway from each work are charted in the Appendix. Also included in the table in the Appendix is an indication (denoted by a circle) of which articles reflected the key themes included in the discussion below, to serve as an aid in situating the themes in the discussion below within the analyzed literature. It is noteworthy that three articles (41–43) identified through the database search were primarily focused on techniques for ERT, but were included as they highlighted key issues around bias, and thus the ethics of ERT. Moreover, multiple included articles were by a research group investigating the implications of ERT, including articles authored or co-authored by McStay (5,9,15,15,18,19,24,36,44–47).

Ethical considerations

The analysis of the literature led to the identification of ethical concerns in three key areas. The first was the risk of biased and unfair outcomes through the use of the systems. This theme encompassed issues which arise as a result of the problematic premises which underpin ERTs, including the shaky science of emotions. The second theme addressed the sensitivity of emotion data which is used in ERTs, including the perceived implications of collecting and utilizing data on human emotion through ERTs. The third theme present in the literature, though overlapping with the first two, was concern over the risk of harm arising from the use of ERT in particular, consequential settings, including workplaces, education, healthcare and policing.

Biases and faulty bases

Multiple studies reported on the risk of bias and unfairness as a key ethical issue in ERT, due at least in part to the problematic premises which underpin it. As described above, there is a notable lack of consensus around what emotion is and how it can be operationalized and measured. ERT are primarily based on the Basic Emotions model, which sees emotion as "distinct natural categories," identifiable from behavior, and generally designed "to only recognize a small number of emotions (e.g., 6) which is hardly representative of real life" (35). This model has been critiqued as "Western-centric" (36), and a product of "Western, Educated, Industrialized, Rich, and Democratic (WEIRD) nations" (16). Perhaps as a result of this limited background, ERT are generally not designed to factor in either immediate or broader contextual factors including culture, despite the relevance of these factors in emotion (4,10,14,16). They thus risk overemphasizing simple facial movements, which is problematic when ERT are used in consequential decision-making, as further discussed below (35). It is noteworthy, however, that Bakir et al. (2,15,36) warn that a shift towards the use of increased contextual data in the form of "multimodal sensing" may lead to even more invasive data collection practices, and would "bake into our urban environments widespread and intensive surveillance of our emotions."

Furthermore, ERT often rely on third-party annotation of data. Hernandez et al. (14) question the basic viability of this approach, arguing that there is "an obvious mismatch between felt emotions, expressed emotions, and perceived emotions." They argue that this is the case not only for third-party annotation, but also for self-report, which can be impacted by "confounds" (14) such as the framing of the question, recall biases, and misattribution of arousal. Moreover, data annotation can introduce biases, as annotation is often based on convenience samples, without sufficient controls, and reflects the implicit biases of those doing the labeling (48). This is particularly problematic when annotation conducted in one setting is later used as the basis for algorithms exported elsewhere; McStay (18) highlights the prevalence of annotation conducted by white Westerners for algorithms applied internationally, though this can also be problematic when the technology is used in a different part of a single country (20).

The biases that occur as a result of these processes are covered by a list compiled by Booth et al. (48) of two broad categories and seven specific types of bias that can arise in machine learning for affective computing broadly. These include, under the category of deficiency: selection/sampling bias and omitted variables; and under the category of contamination: historical, representational, behavioral, presentational, and observer-based biases. The presence of biases based on insufficient representation are increasingly being recognized, with

intersectional identities in particular under-represented in the data used for ERT (41,48). There have been movements towards correcting this bias, especially in relation to gender and racial bias—however, there continues to be a lack of diversity in represented categories such as age, disability, and nationality (41). Yet, bias in ERT is often obscured as it is hidden "under a veneer of scientific objectivity" (1), which makes the systems appear to be neutral.

Furthermore, though they are often used interchangeably, Booth et al. (48) distinguish between bias and unfairness, indicating that bias refers to "any systematic error" which may arise in a system, while fairness is "a subjective perspective" dealing with the appropriateness of the construct itself. As the authors argue, ERT can lead to unfair outcomes, which cannot always be easily remedied (48). These outcomes can include restriction of access to necessary services, manipulation, and violation of human rights in cases where they undermine privacy and autonomy (14). Indeed, in considering stakeholders in ERT, Soper et al. (13) draw attention to a need to focus on disadvantaged users, who may be particularly likely to be exposed to harm. Yet, in the United States, for example, there are restrictions on actively correcting for racial biases used in consequential processes such as hiring due to legal restrictions which require that the systems remain "group unaware" (48).

The sensitivity of emotion data

A second key theme in the literature was the particular sensitivity of data linked to emotion, and the problematization of its use. Ienca and Malgieri (49) propose the concept of "mental data" as a meaningful way to understand the data used for ERT, and to conceptualize its implications. In their definition, mental data is "any data that can be organized and processed to infer the mental states of a person, including their cognitive, affective, and conative states" (49). This data is seen to carry ethical sensitivity as it is intimately tied to what it means to be human, including questions of identity, autonomy, and freedom of thought (2,10,29).

Early research, such as a small study of social media users in the United States conducted by Andalibi and Buss (50), suggests that members of the public share this view of emotion-related data as unique, and different from other personal data in that it can provide "unique insights to behavior and are prone to manipulation; and are intimate, personal, vulnerable, complex and hard to define." This view of emotion as inherently private has been echoed in other research by Grote and Korn (51) and Urquhart et al. (19). Indeed, initial research by McStay (9,24) found that at least half of the studied citizens in the UK found the tracking of emotions in public spaces for advertising purposes to be unacceptable. In this vein, Stark and Hoey (13) argue that emotion data should be considered to be as sensitive as health-related data, though it is rarely treated as such. In fact, ERT were previously primarily used for medical purposes; as commercial applications have expanded, actors are less tightly bounded in their handling and use of data (52). Yet, given that ERT can be used not only for real-time but also for retroactive functions when used to track emotional states over time, the technologies may be inadvertently—or indeed, intentionally—tracking depression, mood disorders, or other states which a data subject may not wish to have known (50,53). Data subjects may be especially vulnerable if this type of data is used by employers, health insurance companies, or advertisers, with heightened risks given that the data subject may not be aware of the tracking they are undergoing, or of the use of their data in this way (19,51,53).

The potential use of data retroactively was reported to be particularly problematic, as it involves the "capture" (24) of emotion and an attendant "loss of ephemerality" (36)—emotions which are experienced as momentary or fleeting are processed and even revisited (24). Linked with this is a process of attempting to make objective phenomena that are inherently subjective and contextual (53). This raises the question of who should be given access to such data, and in particular whether data subjects themselves should be given control over their data (9). In addition, McStay (24) defines the "capture" involved in ERT in a second way, referring to the process of "taking possession by force," elsewhere highlighting the risks that people be viewed as "objects rather than subjects," and as "emotional animals to be biologically mapped and manipulated" (9).

Similarly, Steinert and Friedrich (54) identify four problematic outcomes of the surveillance of emotions: first, a chilling effect on autonomy and authenticity; second, the reinforcement of stereotypes around emotions; third, the risk of "alienation" from one's own emotions; and fourth, strengthened social pressure for one to better control one's own emotions, and thus an increase in emotional burden. Ultimately, Ferraro (7) argues that in light of the United Nations Universal Declaration on Human Rights, individuals can be understood to have inalienable "affective rights," and that it would "constitute a gross breach of ethical design practice to apply such technology and at the same time ignore such widely recognized rights."

The sensitivity of emotion data is heightened in relation to vulnerable populations (7,43,49,55). Though anyone with a form of emotional expression which ERT are not designed to recognize will experience heightened vulnerability to problematic outcomes as a result of their use, three groups were identified in the literature as particularly vulnerable: neurodiverse people, the elderly, and children. As indicated in a report by ARTICLE 19 (14,55), ERT "impose norms about neurotypical behavior on people who do not display it in a way the technology is designed to detect." This is especially of concern if neurodiverse people are underrepresented in training data, while the resulting algorithms are used in consequential decision-making, and if the systems erroneously flag unproblematic patterns of expression as suspicious (19). Similarly, the literature indicates that ERT are not well-suited to the elderly—especially those who may experience an age-related decline in cognition (49), or to children, who tend to express emotions differently from adults (43). Yet, as Ferraro (7) has argued, expecting only a "fully capable individual" to be the subject for ERT would also be problematic, as ascertaining who would fall into this category would involve the disclosure of personal and/or medical information, leading to an "untenable situation."

Use in consequential settings

Extending this consideration of vulnerability, certain studies considered the ethics of applications of ERT in consequential settings: in workplaces, in education, in healthcare, and in policing, as will be discussed below.

The fastest growing area for the implementation of ERT is in workplaces (21). Their uses in employment include use in recruitment processes, boosting productivity, preventing harassment, and for security purposes, with implementation reported at major technology companies such as Amazon and Microsoft (2,21). It is touted as effective in increasing productivity, reducing labor turnover, improving social relationships, and empowering workers (2). It has also been widely used as a part of hiring processes, despite the reported issues of bias (48). Though workplace

applications are still under-researched, early research indicates that the workplace is a highly problematic setting for ensuring that the rights of individuals are respected (14,21,53). In particular, there is concern over the rise and normalization of a new form of exploitative surveillance based on emotions, as an extension of Neo-Taylorism; whereas Taylorism focused on efficiency even at the expense of workers' well-being, Neo-Taylorism sees the well-being of the worker as a key component of profitability (21). For this reason, as Mantello et al. (21) have argued, employers seem to have a stake in workers' emotional states, and tracking their emotional states is viewed as within the purview of employee management. Even when workers are given the opportunity to consent, hierarchical and other pressures in the workplace can impede their free consent, making it difficult to ascertain whether true consent has in fact been obtained (35,49). Ultimately, this is to the detriment of employers, as biometric surveillance has been shown to be negatively correlated with organizational commitment, and to lead to increased technostress and burnout, and reduced dignity, autonomy, and trust (2,21). It is noteworthy here that in Mantello et al.'s (31) study of the acceptability of ERT in the workplace, the authors found acceptance to be related to demographic factors. Namely, a more accepting attitude towards ERT was related to: identifying as a man, being East Asian, having a higher income, having a higher level of education, being non- or less-religious, and being from a collectivist background.

Education is another problematic site for ERT (18,53,55). ARTICLE 19 (55) reports on the use of ERT in education in the Chinese context, and on the resistance to it from administrators, teachers, and students. As McStay (18) has reported, there is a lack of evidence that the use of ERT in educational settings will *not* lead to harm. McStay (18) further identifies several issues in ERT in education, including that it repositions students as "users" of the technology, thus raising particular ethical issues. These include issues around data handling and collection, arguing that the collection of such data represents a mismatch between financial incentives and the wellbeing of students, particularly if the data does not bring direct benefits to students themselves. As McStay (18) argues, the risks to wellbeing include a chilling effect on emotional expression and an overall, general "creepiness" about its use. Indeed, McStay and Rosner (44) have described the ethical sensitivity of the use of ERT in applications related to children more broadly. Regardless, research suggests that education continues to be a site for implementation of ERT (2,4,8,9,13–15,18,20).

Next, Straw (22,53) draws attention to the particular ethical issues of ERT in healthcare, arguing that it threatens patient autonomy, though healthcare had previously been a key site for the implementation of the technologies. Specifically, Straw suggests that ERT brings with it a need to reconsider how confidentiality is understood, as ERT may reveal thoughts or conditions which patients did not intend to disclose. Furthermore, Straw draws attention to the difficult questions of accountability which may arise if a patient is misdiagnosed, or if a condition is not detected in a timely manner through the use of ERT. To this end, Straw suggests that data protection laws for health data should be expanded to better address technologies which—like ERT—are based on health data from other sources. It is noteworthy, however, that Ho et al. (2) indicate that there has been a shift away from the development of ERT for uses in healthcare such as the detection of mental health issues, and towards broader commercial uses.

With attention to yet another consequential setting, Podoletz (20) has documented the implications of the use of ERT in policing, surveillance, and crime management. As Podoletz explains, ERT are used primarily in two ways: for explaining or predicting crime, and for detecting deception, and suggests that an expansion of the use of ERT in these ways can be expected. Yet, Podoletz's review of the literature shows that the capabilities of ERT for these uses remain insufficiently developed, and identifies four areas of concern: accuracy and performance, bias, accountability, and rights- and freedom-related concerns. Given these limitations, Podoletz argues that:

> "…even if emotional AI technologies were to become accurate in revealing thoughts, feelings and intentions, their use in a public urban setting for policing purposes should be resisted in democracies because of the technologies' clash with human rights values and liberties in such societies." (20)

Ensuring ethics?

In light of the ethical issues reported above, multiple studies (5,22,37–39) proposed guidelines, principles, or postulates for more ethical development and implementation of ERT, though none were presented in peer-reviewed journals. The guidelines were synthesized using a qualitative, inductive approach. The results of this synthesis are reported in Table 2, below, with the themes, their respective frequencies across the guidelines sets, and a circle to denote which guideline sets included reference to that particular theme.

Table 2 Results of the analysis of guidelines

Theme	***Frequency***	***Guideline sets***				
		Cowie, 2015	Hernandez et al., 2021	Landowska, 2019	McStay and Pavliscak, 2019	Ong, 2021
Cluster 1: Ethical design						
Ethical decision-making	6	○			○	
Fairness and non-discrimination	6		○		○	○
Privacy	6		○		○	○
Conceptualizing emotion	5		○	○	○	
Quality and validity	4		○	○		○
Ensuring safety	3	○			○	
Cluster 2: Ethical implementation						
Defined scope for use	7	○	○	○	○	○
Transparency	5		○	○		
Consent	4		○		○	
Recognizing limitations	3	○	○			○
Respect	3	○			○	
User control	3		○			
Oversight	2		○			○

There was a total of 57 individual items, in two broad areas which were nearly even: guidelines for ethical design (n=30), and guidelines for ethical implementation (n=27). Within the cluster of ethical design, three themes were equally prominent: ethical decision-making (n=6), fairness and non-discrimination (n=6), and privacy (n=6). In the area of ethical decision-making, Cowie (56) proposed five principles around the need for those designing affective computing systems to be sensitive to the potential for ethical issues, and to act ethically in order to prevent them, calling for developers to be certain that "the systems they build will do nothing to others that they would not wanted to be subjected to themselves." McStay and Pavliscak (45) encouraged ethical review being sought out where there may be "ambiguity" about the ethics of a system. Under the theme of fairness and non-discrimination were guidelines related to the need to have sensitivity to diversity (14,35,45), and to minimize bias (35) by ensuring that the technology had been trained on a diverse dataset (45). In relation to privacy, there were calls to ensure data minimization and privacy as a default through the use of edge processing and/or aggregating data, with attention as well to issues of privacy, consent, and ownership (14,35,45). There was also extension into the area of ethical implementation through the need to recognize that the collection of data in public areas "may be unwanted or invasive" (45).

Conceptualizing emotion (n=5) was a similarly prominent area and pointed to the ongoing debate around the nature of emotion. Principles from Hernandez et al. (14), Landowska (6), and McStay and Pavliscak (45) highlighted the need for a nuanced understanding of emotional expression as just one, outwardly visible component of a more complex phenomenon, which should not be used for prediction, and about the nature of which there is a lack of consensus. A smaller theme was on the need to ensure quality, validity, and the robustness of systems (14,35). One way in which this could be achieved would be through the provision of "guidelines for labeling protocols," to ensure the quality of data used for ERT (6). Finally, the smallest theme in this area was ensuring safety, with Cowie (56) calling for "realistic assessments" of what systems could do and their potential risks—which was echoed by McStay and Pavliscak (45), who additionally called for continency plans to be designed in case a mental health issue were to be detected by a system.

There were seven themes in the area of ethical implementation. The largest was on the need for a defined scope for the use of ERT (n=7), as multiple guidelines called for the delineation of a clear scope for use of ERT, in alignment with a pre-specified purpose (6,35). ERT was seen to be inappropriate for use in making assessments, and should be used only in cases with a clear benefit to the user (45), such as in supporting positive interactions between humans and machines (56). Transparency (n=5) and consent (n=4) were two interrelated themes in this area. The theme of transparency included calls for systems to be described with sufficient detail, and to include indications about "confidence or uncertainty" in relation to the systems (6,14). This was also a prerequisite for consent, which must be free, meaningful, and based on the user's understanding of the system (14,45).

Three smaller themes in this area were the need to recognize the limitations of ERT (n=3), to ensure respect for users (n=3), and to provide avenues for user control (n=3). There must be clarity in the use of ERT about the limitations of the systems, and their predictions should not be treated as "ground truth" (14,35,56). Once implemented, the systems should reflect respect for users and those subject to its decisions (n=3), and should not be deceptive or violate their trust

(45,56). Users should be given control (n=3) over the system through customization options and through the provision of feedback, and data should be under the control of the user with the capability for them to delete their data as needed (14). The final theme in this area was the need for oversight (n=2), by providing a human-in-the-loop and human input, and through monitoring of the "actual outcomes" of the systems, rather than their "intended effects" (14,35).

Yet, despite these attempts to create guidelines for ethical ERT, Urquhart et al. (19) have highlighted issues around attempts to enforce ethics, drawing attention in particular to the challenges in establishing global standards given the diverse contexts in which ERT may be applied. As one part of this, there is a noted disparity in the degree to which ERT is regulated, and many countries have yet to implement sufficient safeguards (5). This facilitates the movement of sites for data collection and implementation into national contexts where "data collection and privacy laws are less stringent" (2). Even in Europe, which has led the charge for greater privacy and control of data through the European General Data Protection Regulation (GDPR), issues remain as the GDPR itself does not directly regulate emotion tracking insofar as it relies on soft, non-identifiable biometric data (5,9,36). However, it is noteworthy that further regulation appears to be on its way in Europe, as the forthcoming AI Act restricts the use of biometric surveillance broadly, and the use of ERT in policing and border management, education, and workplaces, specifically (57).

Moreover, in the absence of appropriate legislation, there is a risk that the pursuit of ethics guidelines contribute to ethics washing, as described by Urquhart et al (19), below:

> "There is concern of notions of 'ethics washing' around AI currently. Yet ethics is still, to a large extent, becoming a branding exercise, much like corporate social responsibility. Companies positioning themselves in the market need to differentiate themselves from competitors and signal their virtues. Crafting their key 'ethical principles' can be a way of doing this. This, in turn, moves away from more accountable norms (like law), towards controlling the terms on which they are judged publicly. Firms may claim to go beyond regulation to build in resilience and a sustainable business that stays ahead of regulatory shifts. Others may use the same rhetoric but not adhere to this in practice." (19)

Wright (58) argues that efforts to increase transparency around the technologies—as proposed in the guidelines above, for example—are insufficient. The author contends that, in light of the "unique privacy and social implications" (53) of ERT, there is a need for greater accountability beyond this, and for the recognition of the broader societal impact of ERT. Ultimately, these initial efforts to promote more ethical implementation of ERT, Stark and Hoey (13,20) argue that safeguards, while important, are insufficient in light of the "potentially toxic social effects," and that:

> "the particularities of how these systems are designed—including the models of emotion designers use to ground their models, and the types of proxy data

> for emotion they collect—matter greatly for the ethical appropriateness of such systems, and even whether they should be developed and deployed at all."

These are sentiments reflected by ARTICLE 19 (55), who call for bans on ERT more broadly.

Discussion

This review has shown that the design and implementation of ERT is being pursued without sufficient recognition of—and despite—its ethical complexity. The corpus of literature identified in this study reports on ethical issues at multiple points across the lifecycle of the technologies, including their fundamental premises, development, and implementation. In so doing, this review has highlighted fundamental issues with ERT in the lack of a scientific consensus around what emotion is. This ties into the question of whether emotion states are phenomena that can be "recognized" based on outward cues, which recent research such as by Barrett et al. (11) has strongly refuted. Moreover, this is further complicated by the involvement of machines in this process. As has been seen, the assumption that emotions *can* be recognized has led to the development of the current techniques used for annotation and development. This includes the use of third-party annotation, alongside self-report—both of which introduce further bias into the annotation process. Bias in ERT is particularly problematic given that AI systems are often assumed to represent a "view from nowhere" (59) and to represent an objective viewpoint; thus, designers, deployers, and users of the technology may not be attuned to these issues. Meanwhile, automation bias may serve to reinforce perceptions of the accuracy of the systems, providing "a veneer of scientific objectivity" (1), while obscuring the unresolved and fundamental issues regarding the unsound "science" (60) behind ERT.

Further to this is the particular sensitivity of attempting to "capture" (24), read, process, and even retroactively reference data on emotions. In this vein, ERT present a particular dilemma: if the technologies can do what they are claimed to, there is the risk that they reveal information about individuals that they may not wish to have known. And, if they cannot, the technologies may be leading to erroneous assumptions about individuals. Given that these issues are related to the viability of the fundamental premises of the technology—and that as the prolonged debate over the nature of emotion has shown, they are potentially irresolvable— they necessarily call into question the entire pursuit of these technologies. Yet, as this review has shown, ERT continue to be applied in highly consequential settings including workplaces, education, healthcare, and law enforcement—each of which has been identified as problematic.

In addition, the studies included in this review have highlighted a lack of sufficient research—and an urgent need for greater research—about the acceptability of these technologies to those who will be subject to them; while even more fundamentally, there is a lack of sufficient awareness of the existence and use of these technologies (41,50,61). Initial research as reported by Mantello et al. (60) suggests unevenness in stakeholder acceptance of the technologies, and a link to particular demographic factors, many of which—including being male, wealthy, and well-educated—are the demographic characteristics of the majority of the individuals in positions of power to decide where and how these technologies are implemented (62,63).

As indicated through the analysis of guidelines included in this study, there are hopes that greater consideration of ethics in both the design and implementation of ERT can be beneficial. Most prominently, this includes the delineation of a pre-defined scope for its use, ensuring that ethical

considerations factor into decision-making, promoting fairness and non-discrimination, and protecting privacy. Given that the risk of bias and unfairness and issues around the use of ERT in consequential settings emerged as key themes in the literature, it was unsurprising to find these issues to be priorities in the guidelines analyzed above. Indeed, as Joyce et al. (64) and Hagendorff (65) have argued, issues such as representativeness and bias are often more easily addressed through technical fixes, and for this reason are often prioritized for resolution. Yet, Munn (66) has argued that these issues—"highly contested" and with "high stakes"—are not so easily resolved, and they raise the question of "who decides" (67)—what counts as fair. Furthermore, a focus on technical fixes can gloss over more fundamental issues inherent in technologies, as argued by Benjamin (68) and Gebru (62). Indeed, as Lauer (69) has argued, this "fallacy of the broken part" may distract from broader, systemic issues and "organization-wide ethical shortcomings," particularly when a focus on ethical decision-making as in the analyzed guidelines locates the source of potential ethical issues such as those related to bias within individuals.

In light of this, it is imperative that the question of proportionality be considered (70), and whether the risks of ERT are truly worth their purported benefits. Though discussions of ethics are often centered around short- and medium-term considerations, there is also a need for consideration of the longer-term implications of these technologies (71). Concern was expressed in this body of literature about some of the ethical implications of ERT; yet, it is noteworthy here that van Wynsberghe (72) has conceptualized attention to AI ethics as occurring in three waves: while a current, second wave has attended to issues of bias, accountability, and transparency, a third, coming wave of ethics, must consider the sustainability of AI itself. The concerns in the included literature and the analyzed guidelines reflect attention to second-wave issues of bias and transparency. However, there has been a gap in the literature in attention to broader sustainability issues, even as Crawford (37), Jaume-Palasi (73), and Brevini (74) have highlighted the environmental toll of the development of AI, the infrastructure for which is in itself under threat from environmental degradation (37,74–77). The costs of AI technologies extend across their lifecycles, with effects that often cross national borders, including through the extraction of materials and the creation of devices and infrastructure, the energy costs and carbon emissions of development and use, and the devastation caused by e-waste (37,74,78). The merits of the technologies must also be considered against the backdrop of the accelerated breach of planetary boundaries, particularly given the stark reality that "[i]f we lose our environment, we lose our planet and our lives" (74). This perspective was largely absent from the literature and from the guidelines analyzed.

Furthermore, a growing body of works in the field of AI ethics documents the challenges of putting well-intentioned ethical principles for AI into practice (79). Munn (66), for example, has argued for the "uselessness" of such ethical principles, due in part to their "toothlessness" (80).These works warn against over-reliance on ethical guidelines to resolve the inherent tensions in AI, and this critique can be applied similarly to ERT. There is, additionally, the risk that such guidelines serve as a kind of "ethics washing" (19,81) which work to appease some critics and deflect attention away from the need for firm legislation and, again, from the fundamental reconsiderations of whether these technologies should be pursued in the first instance (13,58). This is particularly problematic in the case of ERT which, as described above, is used in highly consequential ways, while its most fundamental premises remain questionable

and highly contended. The continued pursuit of these technologies in the face of these issues reflects the profit-driven approach of the commercial interests behind them, and a broader tendency in AI through which, as Elish and boyd (82) write, “spectacle is prioritized over careful consideration of the implications of long-term deployment” (18,19).

There are multiple potential frameworks through which to understand the ethical implications of ERT. Here, a deontological and a consequentialist approach will be considered. A deontological approach would indicate that ERT should not be pursued, at minimum in their current form, given that their use violates fundamental rights to privacy and to integrity (55). Furthermore, as discussed above and as proposed by Ferraro (7), they also violate affective rights through the collection and analysis of intimate data about emotional states. These issues are inherent in the technologies and cannot be easily mitigated given that ERT by definition deal with the capture of emotions. In addition to this, concerns over bias in ERT as described above, and the potential for this to link to real-world discriminatory outcomes and further violation of rights pose additional issues when viewed through a deontological perspective.

A consequentialist approach may highlight the purported benefits of the use of ERT, and the potential to improve quality of life that has been argued for, suggesting that ERT could be used in situations where the benefits outweigh the risks (45). Yet, even here, the unsound science behind the technologies themselves have the potential to obviate these potential benefits. For example, given the lack of consensus about whether emotions can even be detected from external cues, systems to detect emotion such as those intended for use by children with autism may in fact falsely classify the emotional states of individuals, and in this sense be considered misleading and ultimately harmful, In addition, given that the ultimate outcomes of use of ERT may lead to further discrimination or infringement on human rights—such as if the technologies are used in settings where data subjects may be vulnerable, such as in workplaces or schools—a consequentialist approach would also ultimately find the risks to be prohibitive. It is noteworthy that Mantello and Ho (30) argue that an approach based on virtue ethics may be useful in developing appropriate regulatory frameworks, particularly if they drew on diverse perspectives, including “both East and West value traditions, blending the best of Confucian, Buddhist, and Aristotelian virtue ethic traditions.” Yet, as the deontological and consequentialist approaches highlight, regulatory frameworks may be insufficient given the ethical issues posed by these technologies.

Considering this corpus of literature on the ethics of ERT, then, it becomes clear that there remain significant—and potentially irresolvable—ethical issues, not least due to the unsound assumptions the technologies are based on. These findings must call into question the continued commercial development of these technologies for widespread implementation, and highlight an urgent need for stringent oversight of their development and deployment.

Limitations and future directions

ERT are a part of a growing field. For this reason, the scope of the initial retrieval of literature for this study was limited to a focus on articles directly dealing with the topic of the ethics of the technologies, published in the English language. However, this focus on ethics, and the exploratory methodological approach which drew primarily on articles retrievable through the initial database search or identified through reference list searches meant that there may be

relevant literature which was beyond the scope of this study. Moreover, it is noteworthy that though the scope of this study was limited to articles published prior to the first half of 2022, literature focused on the ethics of ERT and related areas continues to emerge, (e.g. (30,83–85), and should be included in future reviews of the field. Future studies may expand on and update this exploratory study through a systematic review of the field, and go beyond literature published in English to understand perceptions of the ethics of ERT in a broader range of settings. It is noteworthy that the methodology utilized in this study included a modified version of thematic analysis, which involves a subjective categorization of the key issues into themes. However, this approach was aligned with the overall exploratory aims of this study.

Furthermore, although literature focused on stakeholder perspectives was not a direct focus of this study, it became clear that there is a significant need for future research in this area, to better understand stakeholder perspectives on the use of ERT, and to add to initial insights furnished by McStay (44) in the UK and by Mantello et al. (60) in Japan. Multiple pioneering works were authored or co-authored by McStay and colleagues, highlighting space for further investigation from diverse perspectives and contexts. Research must be urgently conducted with a range of stakeholders, and with a particular focus on stakeholders who may stand at the intersection of multiple vulnerabilities. Specifically, greater insight is needed into the impact of ERT on neurodiverse people and on other minority and vulnerable groups. Ultimately, in light of the fraught nature of the ethics of ERT, and the rapid investment into its development and implementation regardless of these concerns, greater documentation and evidence for their potential ethical implications is urgently needed.

Statements and declarations

This study was informed by deliberations as a part of collaborative research projects between the Osaka University Research Center on Ethical, Legal, and Social Issues (ELSI Center) and companies including Ricoh and NEC, investigating the ethical, legal, and social issues of emerging technologies. The author gratefully acknowledges the members of the research team at the Osaka University ELSI Center for their input on this study, and especially Professor Atsuo Kishimoto and Dr. Yusuke Shikano for their valuable feedback.
The author does not have financial or proprietary interest in the material discussed in this article.

Appendix

Citation	Aim	Key takeaway	Biases and faulty bases	Sensitivity of emotion data	Consequential settings	Ensuring ethics

Cowie, 2012 (87)	Cooney et al., 2018 (86)	Bryant and Howard, 2019 (43)	Booth et al., 2021 (48)	Bakir et al., 2022 (15)	ARTICLE 19 (55)	Andalibi and Buss, 2020 (50)
To offer a "balanced view" of the ethical issues of emotion-oriented technology, as it is applied at the time of writing.	To identify issues in emotion visualization.	To assess the performance of facial emotion recognition algorithms on images of children.	To investigate bias and fairness in machine learning used for affective computing.	To report on a cross-cultural comparison of the use of emotional AI in Japan and the UK.	To review the ethical implications of the societal use of emotion recognition technologies.	To understand the perspectives of social media users on the use of emotion recognition.
"A high proportion of applications seem ethically neutral. … Many potential negatives apply to technology as a whole. Concerns specifically related to emotion involve creating a lie, by simulate emotions that the systems do not have, or promoting mechanistic conceptions of emotion."	There is a need for careful consideration of potential issues in order to "enable emotion visualization to contribute positively to people's well-being." There is a risk that current technologies "could lead to a world where the richest and most technologically advanced are the only ones who can hide their true emotions."	" [P]opular emotion recognition systems have not thoroughly considered children as a part of their target population. Yet, there is little to no regulation on what categories of people this software can or cannot be used for."	" …[T]here is often a tradeoff between creating the most accurate model possible and reducing bias or enhancing fairness. Model validity, bias, and fairness are all crucial considerations for automated AC systems."	" Emotional AI, and wider automated human-state measurement, thus requires ongoing social, cultural, legal, and ethical scrutiny if respect and dignity are to be served."	" Our report demonstrates the need for strategic and well-informed advocacy against the design, development, sale, and use of emotion recognition technologies. We emphasise that the timing of such advocacy – before these technologies become widespread – is crucial for the effective promotion and protection of people' s rights, including their freedoms to express and opine."	" The majority of participants were uncomfortable with emotion recognition, and this discomfort was often related to concerns over privacy, consent, agency, and potential harm."
○	○	○	○	○	○	○
○	○	○	○	○	○	○
	○	○	○	○	○	○
○					○	

Greene, 2020 (4)	Glenn and Monteith, 2014 (52)	Ghotbi et al., 2022(88)	Ghotbi and Ho, 2021(61)	Ferraro, 2020 (7)	Crawford, 2021 (1)	Cowie, 2015 (56)
To provide a basis for multi-stakeholder interaction on affective computing.	To discuss the impact on psychiatry of information about mental states and behaviors collected through emerging technologies.	To identify the ethical issues related to AI of most concern to Japanese college students.	To identify student concerns about the use of artificial intelligence in the future.	To propose affective rights as a new concept, with a prototype for a "Universal Individual Rights Manager."	To present a perspective on emotion recognition in society.	To map the ethical principles underlying debate about ethics in affective computing.
“ We should think about which questions will help us the most in exploring the ethical issues and unintended consequences of creating and deploying AI connected to emotion and affect. We should direct our ethical analysis to the applications that are now coming into the marketplace and that are just behind them in the research labs.”	“ There are many technical issues to resolve, primarily relating to reliability, usability, privacy, and clinical utility of the new measures. At the same time, the use of non-medical commercial products based on similar measures will become pervasive. Society must address the ethical issues associated with measures of mental state and behavior for medical and non-medical use.”	"The majority of students (n=149, 65%) chose unemployment as the major ethical issue related to AI. The second largest group of students (n=29, 13%) were concerned with ethical issues related to emotional AI, including the impact of AI on human behavior and emotion. The paper … concludes that, while policymakers must consider how to ameliorate the impact of AI on employment, AI engineers need to consider the emotional aspects of AI in research and development, as well."	“ The results show that most of the students (n=269, 58%) considered unemployment to be the major ethical issue related to AI. The second largest group of students (n=54, 12%) was concerned with ethical issues related to emotional AI, including the impact of AI on human behavior and emotion and robots' rights and emotions.”	“ Changes will be required in the way consumers, developers, and regulatory agenceis protect [affective] rights, and are made aware of these issues. To that it end, it is proposed independent auditing services be created as part of the prototype to verify and validate stakeholder compliance to accepted ethical standards.”	"There is deep scientific disagreement about whether AI can detect emotions" and "[g]rowing scientific concern about the use and misuse of these technologies.”	There are eight “ ethical obligations that anyone who worked in affective computing should respect.”
○		○	○	○	○	○
○	○	○		○	○	○
○	○			○	○	○
				○		○

Ienca and Malgieri, 2022 (49)	Ho et al., 2021(2)	Hernandez et al., 2021 (14)	Grote and Korn, 2017 (51)	Grond et al., 2019 (89)
To propose the concept of "mental data" and analyze existing legal protections, with a focus on the European context.	To identify the national actors and relationships between them in the area of affective computing.	To identify current applications of emotion recognition, and to propose guidelines for identifying and offsetting risks.	To identify insufficiencies relevant to affective computing in the Code of Ethics for the Association for Computing Machinery.	To consider biomusic for use by individuals on the autism spectrum.
The European General Data Protection Regulation " is an adequate tool to mitigate risks related to mental data processing," but there is a need to consider not only the type of data but also how it will be processed. To this end, a " specific data protection impact assessment" is proposed.	" Contrary to the ongoing political rhetoric of a new Cold War, we argue that there are in fact vibrant AI research alliances and ongoing collaborations between the West and China, especially with the US, despite competing interests and ethical concerns. Our analysis also uncovers a major shift in the focus of affective computing research away from diagnosis and detection of mental illnesses to more commercially viable applications in smart city design."	" As emotion recognition tools become more ubiquitous, there is potential for misuse and harm. This work reviews the current landscape of emotion recognition applications, as well as some of the most common types of harm, such as denial of consequential services, physical and emotional injury, and infringement on human rights. Some of the central challenges include that emotion recognition is based on theories that are still evolving, that the ground truth of emotions is not easily quantifiable, and that large individual and cultural differences exist in terms of emotional experience and expression. Further, the language used to communicate about emotion recognition technologies often over-promises what it can be used accurately for, and rarely reflects the possibility of potential misuse and automation bias."	" Emotion recognition, especially from a distance and without body sensors, endangers central concepts of human autonomy and privacy. From both an ethical and an HCI [human-computer interaction] perspective, we discuss how the ACM code of ethics (CODE) could be revised to better address such challenges. We suggest to employ positive law and to use ' middle-range-principle' to build a substructure of the CODE that researchers in HCI can apply more easily."	" Designing affective technologies necessarily encroaches on the intimate realm of emotions as well as the personal values and sensitivities intertwined with the expression of emotions. While participation is therefore often an uncomfortable process, the inclusion of all stakeholders is critical to ensure that the ethical issues associated with SIIF-based [semi-intelligent information filters] affective technologies are considered during the creation of these technologies."
○	○	○		○
○	○	○	○	○
○	○	○		
○	○	○	○	

McStay, 2018 (24)	McStay and Urquhart, 2019 (36)	McStay and Rosner, 2021 (44)	McStay and Pavliscak, 2019 (45)	Mantello et al., 2021 (21)	Landowska, 2019 (6)	Kim et al., 2021 (41)
To explore the implications of a rise of " empathic media" in the form of emotional AI.	To examine the legal and privacy issues of "appraisal based emotion capture."	To explore the acceptability and governance of emotional AI used in devices used by children.	To propose a checklist for the ethical use of emotional AI.	To explore attitudes towards empathic surveillance.	To identify the implications of "uncertainty" in emotion recognition technologies.	To assess the performance of facial remotion recognition algorithms on images of older adults.
" …while all might enjoy and appreciate the focus on ' experience' (user, consumer, patient and citizen), it is paramount that people have meaningful choice and control over the ' capturing' of information about emotion and their bodies."	"This paper has argued that we need to prepare for the emergence of appraisal-based emotional AI (EAI). It first briefly depicted the historical context of face-based EAI, highlighted weaknesses in 'basic' approaches, and then argued that wide recognition foliitations with these methods will inevitably create interest in appraisal-based approaches. These will be more invasive due to need for extra data on internal (involving metabolic and experiential factors) and external contexts (such as when and where a person is, who they are with, perhaps even what they are saying, and by what means, e.g., in -person or through a device."	"The article highlights disquiet about the evolution of generational unfairness, that encompasses injustices regarding the datafication of childhood, manipulation, parental vulnerability, synthetic personalities, child and parental media literacy, and need for improved governance."	A 15-guideline checklist, with items related to personal, relational, and societal considerations in relation to emotional AI.	" …[A]ffect tools, left unregulated in the workplace, may lead to heightened stress and anxiety among disadvantaged ethnicities, gender and income class…[and] may create more problems in terms of algorithmic bias, opaque decisionism, and the erosion of employment relations."	"[T]here is uncertainty inherent in emotion recognition technologies, and the phenomenon is not expressed enough, not addressed enough and unknown by the users of the technology."	"Our results found that all four commercial [facial emotion recognition] systems most accurately perceived emotion in images of young adults and least accurately in images of older adults."
○	○	○		○	○	○
○	○	○		○	○	○
○	○	○		○	○	○
○	○	○	○	○	○	○

Ong, 2021 (35)	McStay, 2021 (47)	McStay, 2020 (46)	McStay, 2019 (9)	McStay, 2019 (5)	McStay, 2020 (18)
To propose guidelines for evaluating the ethical and moral implications of "affectively-aware AI."	To add to "what are broadly individualistically and Western-oriented ethical debates" about emotional AI from a Japanese perspective.	To identify approaches to the privacy aspects of "soft non-identifying emotional AI."	To offer recommendations to the United Nations Office of the High Commissioner for Human Rights on the use of emotional AI.	To identify the issues with emotional AI, with a focus on advertising.	To evaluate the use of emotional AI in education.
“ We propose a multi-stakeholder analysis framework that separates the ethical responsibilities of AI developers vis-à-vis the entities that deploy such AI--which we term Operators. Our analysis produces two pillars that clarify the responsibilities of each of these stakeholders: Provable Beneficence, which rests on proving the effectiveness of the AI, and Responsible Stewardship, which governs responsible collection, use, and storage of data and the decisions made from such data."	Ethical principles based on the Japanese concepts of community, wholeness, sincerity, and heart are viable.	There is “ a weak consensus among social stakeholders on the need for privacy, this driven by different interests and motivations. … [T]here exists a limited window of opportunity to societally agree principles of practice regarding privacy and the use of data about emotions."	Three key recommendations were identified: “ 1. The OHCHR and Special Rapporteur on the right to privacy should reflect on the social desirability of ‘ machine-readable’ emotional life. 2. While there is certainly scope to connect information about emotions with personal data, urgent attention should be paid to practices that passively read expressions and emotional behaviour. 3. Recognition should be made of the right to ‘ community privacy’ even when individuals are not singled-out.”	“ There is nothing innately wrong with technologies that function in relation to human affect states and emotion. They can serve, assist and entertain, if they are built and deployed in a way that respects the wishes of individuals and groups. The reality however is that this would be a volte-face from the compliance-based approach currently taken by the 'big tech' industry that will increasingly deploy these novel modes of profiling and human-agent interaction.”	There is a “ clash of private and public interests. Concern is … two-fold: first is on method, especially given scope for material effects on students; second are ethical and legal concerns.”
○	○	○	○		○
○	○	○	○		○
○			○	○	○
○	○	○	○	○	○

Stark and Hoey, 2021 (13)	Soper et al., 2020 (8)	Smith et al., 2021 (3)	Shimo, 2020 (16)	Sham et al., 2022 (42)	Sedenberg and Chuang, 2017 (53)	Podoletz, 2022 (20)
To propose models and data for the analysis of emotional expression in humans.	To "standardize the ethical design" of empathetic technologies.	To review the literature on uses of affective computing in the area of mood and cognitive disorders in later life.	To examine ethical concerns about diversity in AI systems, with a focus on emotion recognition technologies.	To investigate racial bias in commonly used emotion recognition methods.	To identify the privacy and policy implications of emotion AI.	To identify the issues related to the use of emotional AI in the areas of criminology, policing, and surveillance.
“ We argue that we should not take computer scientists at their word that the paradigms for human emotions they have developed internally and adapted from other disciplines can produce ground truth about human emotions; instead, we ask how different conceptualizations of what emotions are, and how they can be sensed, measured and transformed into data, shape the ethical and social implications of these AI systems.”	"It is clear from the development of use cases that serious care must be taken in interpreting emotion across culture and gender. The approach being used by businesses today is trial and error and training and learning from past experiences. The key to the future is to create standards and best practices that improve the ethical issues ahead of us.”	There is a need"[t]o optimize the utility of affective computing while mitigating potential risks” and to “ ensure responsible development, ethical development of affective computing applications for late-life mood and cognitivie disorders…”	“ It is suggested that the AI sector re-evaluate diversity, equity, and inclusion practices in their workplace and conduct more inclusive meta-analyses to build balanced emotion AI systems.”	The methods evaluated are more effective for races included in the training data. The authors propose “ adding the missing races into the training data equally."	“ Emotion analysis has great potential to add on to existing digital infrastructure with ease and result in a variety of benefits and risks. Our research points to the unique privacy and social implications of emotion AI technology and the impact it may have on both communities and individuals.”	“ I argue that these technologies should not be deployed within public spaces because there is only a very weak evidence-base as to their effectiveness in a policing and security context, and even more importantly represent a major intrusion to people's private lives and also represent a worrying extension of policing power because of the possibility that intentions and attitudes may be inferred.”
○	○	○	○	○	○	○
○	○	○	○	○	○	○
○	○	○		○	○	○
○	○	○	○		○	○

Steinert and Friedrich, 2019 (54)	To consider the ethical aspects of affective brain-computer interfaces (BCIs) in recent use.	“ Although the development of affective BCIs is still at an early stage, concrete ethical issues can already be identified and should be discussed.”	○	○	○	
Straw, 2021 (22)	To provide an overview of the ethical issues involved in the use of emotion-related AI in mental health.	“ Despite the growth of these technologies, there has not been a parallel growth in the ethical debate. For digital health to be implemented in an equitable manner, clinicians must be ethically equipped to appraise these systems.”	○	○	○	○
Urquhart et al., 2019 (19)	To report the results of Japanese and UK cross-cultural workshops on emotional AI.	Differences between Japanese and UK perspectives on emotional AI were identified, and the importance of cultural context was highlighted.	○	○	○	○
Wright, 2021 (58)	To report what is known about Vibraimage, and how it has come to be widely implemented.	Vibraimage is a "suspect" technology, which lacks appropriate scientific grounding and is rooted in unethical data.	○	○	○	○

References

1. Crawford K. Time to regulate AI that interprets human emotions. Nature. 2021;592.

2. Ho MT, Mantello P, Nguyen HKT, Vuong QH. Affective computing scholarship and the rise of China: a view from 25 years of bibliometric data. Humanit Soc Sci Commun. 2021 Dec;8(1):282.

3. Smith E, Storch EA, Vahia I, Wong STC, Lavretsky H, Cummings JL, et al. Affective Computing for Late-Life Mood and Cognitive Disorders. Vol. 12, FRONTIERS IN PSYCHIATRY. AVENUE DU TRIBUNAL FEDERAL 34, LAUSANNE, CH-1015, SWITZERLAND: FRONTIERS MEDIA SA; 2021.

4. Greene G. The Ethics of AI and Emotional Intelligence: Data sources, applications, and questions for evaluating ethics risk. Partnership on AI; 2020.

5. McStay A. Emotional AI: A societal challenge. KENNISCENTRUM DATA Maatsch. 2020;

6. Landowska A. Uncertainty in emotion recognition. Vol. 17, JOURNAL OF INFORMATION COMMUNICATION & ETHICS IN SOCIETY. HOWARD HOUSE, WAGON LANE, BINGLEY BD16 1WA, W YORKSHIRE, ENGLAND: EMERALD GROUP PUBLISHING LTD; 2019. p. 273–91.

7. Ferraro A. Affective Rights: A Foundation for Ethical Standards. In: 2020 IEEE International Symposium on Technology and Society (ISTAS) [Internet]. Tempe, AZ, USA: IEEE; 2020 [cited 2022 Jun 23]. p. 1–11. Available from: https://ieeexplore.ieee.org/document/9462172/

8. Soper R, Bennet K, Rivas P, Mathana. Developing Use Cases to Support an Empathic Technology Ethics Standard. In: 2020 IEEE International Symposium on Technology and Society (ISTAS) [Internet]. Tempe, AZ, USA: IEEE; 2020 [cited 2022 Jun 23]. p. 25–8. Available from: https://ieeexplore.ieee.org/document/9462177/

9. McStay A. The right to privacy in the age of emotional AI [Internet]. 2019 [cited 2022 Nov 18]. Available from: https://www.ohchr.org/sites/default/files/Documents/Issues/DigitalAge/ReportPrivacyinDigitalAge/AndrewMcStayProfessor_of_Digital_Life,_BangorUniversityWalesUK.pdf

10. Stark L. Facial Recognition is the Plutonium of AI. XRDS. 2019 Apr;25(3):50–5.

11. Barrett LF, Adolphs R, Marsella S, Martinez AM, Pollak SD. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. Psychol Sci. 2019;70.

12. Andrejevic M, Selwyn N. Facial Recognition. Cambridge: Polity Press; 2022.

13. Stark L, Hoey J. The Ethics of Emotion in Artificial Intelligence Systems. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency [Internet]. Virtual Event Canada: ACM; 2021 [cited 2022 Jul 5]. p. 782–93. Available from: https://dl.acm.org/doi/10.1145/3442188.3445939

14. Hernandez J, Lovejoy J, McDuff D, Suh J, O'Brien T, Sethumadhavan A, et al. Guidelines for Assessing and Minimizing Risks of Emotion Recognition Applications. 2021 9TH INTERNATIONAL CONFERENCE ON AFFECTIVE COMPUTING AND INTELLIGENT INTERACTION (ACII). 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE; 2021. (International Conference on Affective Computing and Intelligent Interaction).

15. Bakir V, Ghotbi N, Ho TM, Laffer A, Mantello P, McStay A, et al. Emotional AI in cities: Cross-cultural lessons from the UK and Japan on designing for an ethical life. In: Machine Learning and the City: Applications in Architecture and Urban Design. John Wiley & Sons Ltd; 2022.

16. Shimo S. Risks of Bias in AI-Based Emotional Analysis Technology from Diversity Perspectives. In: 2020 IEEE International Symposium on Technology and Society (ISTAS) [Internet]. Tempe, AZ, USA: IEEE; 2020 [cited 2022 Jun 23]. p. 66–8. Available from: https://ieeexplore.ieee.org/document/9462168/

17. Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine [Internet]. New York, NY: Springer New York; 2013 [cited 2022 Jul 10]. Available from: http://link.springer.com/10.1007/978-1-4419-1005-9

18. McStay A. Emotional AI and EdTech: serving the public good? Vol. 45, LEARNING MEDIA AND TECHNOLOGY. 2-4 PARK SQUARE, MILTON PARK, ABINGDON OX14 4RN, OXON, ENGLAND: ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD; 2020. p. 270–83.

19. Urquhart L, McStay A, Mantello P, Bakir V. Emotional AI: Japan and UK Final Report on a Conversation Between Cultures. 2019.

20. Podoletz L. We have to talk about emotional AI and crime. AI Soc [Internet]. 2022 May 5 [cited 2022 Nov 17]; Available from: https://link.springer.com/10.1007/s00146-022-01435-w

21. Mantello P, Ho MT, Nguyen MH, Vuong QH. Bosses without a heart: socio-demographic and cross-cultural determinants of attitude toward Emotional AI in the workplace. AI Soc [Internet]. 2021 Nov 6 [cited 2022 Jun 23]; Available from: https://link.springer.com/10.1007/s00146-021-01290-1

22. Straw I. Ethical implications of emotion mining in medicine. Health Policy Technol. 2021 Mar;10(1):191–5.

23. Garcia-Garcia JM, Penichet VMR, Lozano MD, Fernando A. Using emotion recognition technologies to teach children with autism spectrum disorder how to identify and express emotions. Univers Access Inf Soc. 2022 Nov;21(4):809–25.

24. McStay A. Emotional AI: The rise of empathic media. London: SAGE Publications; 2018.

25. Subramanian B, Kim J, Maray M, Paul A. Digital Twin Model: A Real-Time Emotion Recognition System for Personalized Healthcare. IEEE Access. 2022;10:81155–65.

26. Talaat FM. Real-time facial emotion recognition system among children with autism based on deep learning and IoT. Neural Comput Appl. 2023 Jun;35(17):12717–28.

27. Kalantarian H, Jedoui K, Washington P, Tariq Q, Dunlap K, Schwartz J, et al. Labeling images with facial emotion and the potential for pediatric healthcare. Artif Intell Med. 2019 Jul;98:77–86.

28. Blocher K, Picard RW. Affective Social Quest: Emotion Recognition Therapy for Autistic Children. In: Dautenhahn K, Bond A, Cañamero L, Edmonds B, editors. Socially Intelligent Agents [Internet]. Boston, MA: Springer US; 2002 [cited 2023 May 22]. p. 133–40. (Weiss G, Carley KM, Demazeau Y, Durfee E, Gasser L, Gilbert N, et al., editors. Multiagent Systems, Artificial Societies, and Simulated Organizations; vol. 3). Available from: http://link.springer.com/10.1007/0-306-47373-9_16

29. Pavez R, Diaz J, Arango-Lopez J, Ahumada D, Mendez-Sandoval C, Moreira F. Emo-mirror: a proposal to support emotion recognition in children with autism spectrum disorders. Neural Comput Appl. 2023 Apr;35(11):7913–24.

30. Mantello P, Ho MT. Emotional AI and the future of wellbeing in the post-pandemic workplace. AI Soc [Internet]. 2023 Feb 7 [cited 2023 May 21]; Available from: https://link.springer.com/10.1007/s00146-023-01639-8

31. Henkel AP, Bromuri S, Iren D, Urovi V. Half human, half machine – augmenting service employees with AI for interpersonal emotion regulation. J Serv Manag. 2020 Jul 7;31(2):247–65.

32. Zepf S, Hernandez J, Schmitt A, Minker W, Picard RW. Driver Emotion Recognition for Intelligent Vehicles: A Survey. ACM Comput Surv. 2021 May 31;53(3):1–30.

33. Boyd KL, Andalibi N. Automated Emotion Recognition in the Workplace: How Proposed Technologies Reveal Potential Futures of Work. Proc ACM Hum-Comput Interact. 2023 Apr 14;7(CSCW1):1–37.

34. Levy K. Data driven: truckers, technology, and the new workplace surveillance. Princeton University Press; 2022.

35. Ong DC. An Ethical Framework for Guiding the Development of Affectively-Aware Artificial Intelligence. 2021 9TH INTERNATIONAL CONFERENCE ON AFFECTIVE COMPUTING AND INTELLIGENT INTERACTION (ACII). 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE; 2021. (International Conference on Affective Computing and Intelligent Interaction).

36. McStay A, Urquhart, L. "This time with feeling?" Assessing EU data governance implications of out of home appraisal based emotional AI. First Monday. 2019;

37. Crawford K. Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. New Haven: Yale University Press; 2021.

38. Mohammad SM. Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis [Internet]. arXiv; 2022 [cited 2022 Oct 3]. Available from: http://arxiv.org/abs/2109.08256

39. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006 Jan;3(2):77–101.

40. Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. SSRN Electron J [Internet]. 2020 [cited 2022 Dec 20]; Available from: https://www.ssrn.com/abstract=3518482

41. Kim E, Bryant D, Srikanth D, Howard A. Age Bias in Emotion Detection: An Analysis of Facial Emotion Recognition Performance on Young, Middle-Aged, and Older Adults. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society [Internet].

Virtual Event USA: ACM; 2021 [cited 2022 Jun 23]. p. 638–44. Available from: https://dl.acm.org/doi/10.1145/3461702.3462609

42. Sham AH, Aktas K, Rizhinashvili D, Kuklianov D, Alisinanoglu F, Ofodile I, et al. Ethical AI in facial expression analysis: racial bias. Signal Image Video Process [Internet]. 2022 May 9 [cited 2022 Jun 23]; Available from: https://link.springer.com/10.1007/s11760-022-02246-8

43. Bryant D, Howard A. A Comparative Analysis of Emotion-Detecting Al Systems with Respect to Algorithm Performance and Dataset Diversity. AIES `19: PROCEEDINGS OF THE 2019 AAAI/ACM CONFERENCE ON AI, ETHICS, AND SOCIETY. 1515 BROADWAY, NEW YORK, NY 10036-9998 USA: ASSOC COMPUTING MACHINERY; 2019. p. 377–82.

44. McStay A, Rosner G. Emotional artificial intelligence in children's toys and devices: Ethics, governance and practical remedies. Big Data Soc. 2021 Jan;8(1):205395172199487.

45. McStay A, Pavliscak P. Emotional Artificial Intelligence: Guidelines for ethical use [Internet]. 2019 [cited 2022 Oct 24]. Available from: https://drive.google.com/file/d/1frAGcvCY_v25V8ylqgPF2brTK9UVj_5Z/view

46. McStay A. Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. Big Data Soc. 2020 Jan;7(1):205395172090438.

47. McStay A. Emotional AI, Ethics, and Japanese Spice: Contributing Community, Wholeness, Sincerity, and Heart. Philos Technol. 2021 Dec;34(4):1781–802.

48. Booth BM, Hickman L, Subburaj SK, Tay L, Woo SE, D'Mello SK. Integrating Psychometrics and Computing Perspectives on Bias and Fairness in Affective Computing: A case study of automated video interviews. IEEE Signal Process Mag. 2021 Nov;38(6):84–95.

49. Ienca M, Malgieri G. Mental data protection and the GDPR. Vol. 9, JOURNAL OF LAW AND THE BIOSCIENCES. GREAT CLARENDON ST, OXFORD OX2 6DP, ENGLAND: OXFORD UNIV PRESS; 2022.

50. Andalibi N, Buss J. The Human in Emotion Recognition on Social Media: Attitudes, Outcomes, Risks. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems [Internet]. Honolulu HI USA: ACM; 2020 [cited 2022 Jun 23]. p. 1–16. Available from: https://dl.acm.org/doi/10.1145/3313831.3376680

51. Grote T, Korn O. Risks and Potentials of Affective Computing. Why the ACM Code of Ethics Requires a Substantial Revision. 2017;6.

52. Glenn T, Monteith S. New Measures of Mental State and Behavior Based on Data Collected From Sensors, Smartphones, and the Internet. Vol. 16, CURRENT PSYCHIATRY REPORTS. 233 SPRING ST, NEW YORK, NY 10013 USA: SPRINGER; 2014.

53. Sedenberg E, Chuang J. Smile for the Camera: Privacy and Policy Implications of Emotion AI. arXiv. 2017;

54. Steinert S, Friedrich O. Wired Emotions: Ethical Issues of Affective Brain-Computer Interfaces. Vol. 26, SCIENCE AND ENGINEERING ETHICS. VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS: SPRINGER; 2020. p. 351–67.

55. ARTICLE 19. Emotional Entanglement: China's emotion recognition market and its implications for human rights [Internet]. 2021 [cited 2022 Oct 24]. Available from: https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf

56. Cowie R. Ethical Issues in Affective Computing. In: Calvo R, D'Mello S, Gratch J, Kappas A, editors. The Oxford Handbook of Affective Computing [Internet]. Oxford University Press; 2015 [cited 2022 Oct 5]. p. 0. Available from: https://doi.org/10.1093/oxfordhb/9780199942237.013.006

57. European Parliament. AI Act: a step closer to the first rules on Artificial Intelligence. 2023.

58. Wright J. Suspect AI: Vibraimage, Emotion Recognition Technology, and Algorithmic Opacity. SSRN Electron J [Internet]. 2020 [cited 2022 Aug 31]; Available from: https://www.ssrn.com/abstract=3682874

59. Katz Y. Artificial whiteness: politics and ideology in artificial intelligence. New York: Columbia University Press; 2020.

60. Mantello P, Ho MT, Nguyen MH, Vuong QH. Bosses without a heart: socio-demographic and cross-cultural determinants of attitude toward Emotional AI in the workplace. AI & SOCIETY. ONE NEW YORK PLAZA, SUITE 4600, NEW YORK, NY, UNITED STATES: SPRINGER;

61. Ghotbi N, Ho MT. Moral Awareness of College Students Regarding Artificial Intelligence. Asian Bioeth Rev. 2021 Dec;13(4):421–33.

62. Gebru T. Race and Gender. In: Dubber MD, Pasquale F, Das S, editors. The Oxford Handbook of Ethics of AI [Internet]. 2020 [cited 2022 Nov 4]. p. 252–69. Available from: https://academic.oup.com/edited-volume/34287/chapter-abstract/290662826?redirectedFrom=fulltext

63. Abbate J. Coding is not empowerment. In: Mullaney TS, Peters B, Hicks M, Philip K, editors. Your computer is on fire. Cambridge: The MIT PRess; 2021.

64. Joyce K, Smith-Doerr L, Alegria S, Bell S, Cruz T, Hoffman SG, et al. Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change. Socius Sociol Res Dyn World. 2021 Jan;7:237802312199958.

65. Hagendorff T. The Ethics of AI Ethics: An Evaluation of Guidelines. Minds Mach. 2020 Mar;30(1):99–120.

66. Munn L. The uselessness of AI ethics. AI Ethics [Internet]. 2022 Aug 23 [cited 2022 Nov 4]; Available from: https://link.springer.com/10.1007/s43681-022-00209-w

67. Zuboff S. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. London: Public Affairs; 2019.

68. Benjamin R. Race After Technology. Vol. 2019. Cambridge: Polity Press;

69. Lauer D. You cannot have AI ethics without ethics. AI Ethics. 2021 Feb;1(1):21–5.

70. Karliuk M. Proportionality principle for the ethics of artificial intelligence. AI Ethics [Internet]. 2022 Oct 6 [cited 2022 Dec 6]; Available from: https://link.springer.com/10.1007/s43681-022-00220-1

71. Ryan M, Antoniou J, Brooks L, Jiya T, Macnish K, Stahl B. Research and Practice of AI Ethics: A Case Study Approach Juxtaposing Academic Discourse with Organisational Reality. Sci Eng Ethics. 2021 Apr;27(2):16.

72. van Wynsberghe A. Sustainable AI: AI for sustainability and the sustainability of AI. AI Ethics. 2021 Aug;1(3):213–8.

73. Jaume-Palasi L. Why We Are Failing to Understand the Societal Impact of Artificial Intelligence. Soc Res. 2019;86(2):477–98.

74. Brevini B. Is AI good for the planet? Cambridge: Polity Press; 2021.

75. Stokel-Walker C. Data centers are facing a climate crisis [Internet]. 2022 [cited 2022 Nov 30]. Available from: https://www.wired.co.uk/article/data-centers-climate-change

76. Dauvergne P. AI in the wild: Sustainability in the age of artificial intelligence. Cambridge: The MIT Press; 2020.

77. Wallace-Wells D. The Uninhabitable Earth. New York: Tim Duggan Books; 2019.

78. Parvez SM, Jahan F, Brune MN, Gorman JF, Rahman MJ, Carpenter D, et al. Health consequences of exposure to e-waste: an updated systematic review. Lancet Planet Health. 2021 Dec;5(12):e905–20.

79. Floridi L. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. Philos Technol. 2019 Jun;32(2):185–93.

80. Rességuier A, Rodrigues R. *AI ethics should not remain toothless!* A call to bring back the teeth of ethics. Big Data Soc. 2020 Jul;7(2):205395172094254.

81. van Maanen G. AI Ethics, Ethics Washing, and the Need to Politicize Data Ethics. Digit Soc. 2022 Sep;1(2):9.

82. Elish MC, boyd danah. Situating methods in the magic of Big Data and AI. Commun Monogr. 2018 Jan 2;85(1):57–80.

83. Mantello P, Ho MT. Why we need to be weary of emotional AI. AI Soc. 2022 Oct 11;s00146-022-01576-y.

84. Gremsl T, Hödl E. Emotional AI: Legal and ethical challenges1. Čas J, De Hert P, Porcedda MG, Raab CD, editors. Inf Polity. 2022 Jul 26;27(2):163–74.

85. Ghotbi N. The Ethics of Emotional Artificial Intelligence: A Mixed Method Analysis. Asian Bioeth Rev [Internet]. 2022 Dec 2 [cited 2022 Dec 8]; Available from: https://link.springer.com/10.1007/s41649-022-00237-y

86. Cooney M, Pashami S, Sant'Anna A, Fan Y, Nowaczyk S. Pitfalls of Affective Computing How can the automatic visual communication of emotions lead to harm, and what can be done to mitigate such risks? COMPANION PROCEEDINGS OF THE WORLD WIDE WEB CONFERENCE 2018 (WWW 2018). 1515 BROADWAY, NEW YORK, NY 10036-9998 USA: ASSOC COMPUTING MACHINERY; 2018. p. 1563–6.

87. Cowie R. The Good Our Field Can Hope to Do, the Harm It Should Avoid. Vol. 3, IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. 445 HOES LANE, PISCATAWAY, NJ 08855-4141 USA: IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC; 2012. p. 410–23.

88. Ghotbi N, Ho MT, Mantello P. Attitude of college students towards ethical issues of artificial intelligence in an international university in Japan. AI Soc. 2022 Mar;37(1):283–90.

89. Grond F, Motta-Ochoa R, Miyake N, Tembeck T, Park M, Blain-Moraes S. Participatory Design of Affective Technology: Interfacing Biomusic and Autism. Vol. 13, IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. 445 HOES LANE, PISCATAWAY, NJ 08855-4141 USA: IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC; 2022. p. 250–61.