Ethical Considerations of Facial Expression Recognition AI for Human-Robot Interactions

De'Aira Bryant[1] and Ayanna Howard[2]

Abstract— **This paper explores the ethical considerations surrounding Facial Expression Recognition (FER) AI in Human-Robot Interactions (HRI), focusing on whether and how robots should perceive and interpret human facial expressions. It examines the implications for privacy, user consent, and societal integration, applying existing frameworks and proposing four ethical approaches: Ethical Non-Use, Visible Cue Perception, Necessary Informed Consent, and Contextual Appropriateness. A privacy risk matrix is introduced to evaluate these approaches, highlighting potential risks such as invasion of privacy, algorithmic bias, data misuse, and consent mismanagement. The paper underscores the need for proactive measures in AI development, including auditing, bias mitigation, and contextually sensitive safeguards, to ensure responsible deployment of FER technology. By addressing these ethical dimensions, the paper contributes to advancing a future where AI technologies in robotics are aligned with ethical principles, promoting fairness, transparency, and user trust.**

I. INTRODUCTION

Autonomous agents that are situated in social environments, oftentimes referred to as social robots or embodied agents, commonly use perceptual systems that support facial analysis to better understand and engage with human interactants. While they have been increasingly adopted for applications in various domains like education and healthcare, research in innovation adoption suggests that socially autonomous agents will need to be technically and emotionally sophisticated to capture the interest of mass markets [38]. Facial expression recognition (FER) AI systems have the potential to help bridge this gap and have been considered in some human-robot interaction (HRI) research [16], [30]. In the field of education, this technology may empower agents to support teachers in monitoring students' moods and engagement levels [15], [20]. In the medical field, doctors may utilize the technology to aid in the diagnosis, monitoring, and treatment of mental and psychiatric conditions [41], [6]. There have also been proposals where the technology could enhance driver safety within the automotive industry by assisting in the detection of fatigue or drowsiness [44], [20]. It could furthermore enhance personalization for products tailored to emotional states, such as music playlists and product recommendations, in the entertainment and marketing industries [13]. Advocates for the technology suggest that an optimal system may fulfill these roles in a manner that is more objective and scientifically grounded than typical human judgment [20].

[1]De'Aira Bryant is with the College of Computing, Georgia Institute of Technology, Atlanta, GA 30308, USA dbryant@gatech.edu

[2]Ayanna Howard is with the College of Engineering, Ohio State University, Columbus, OH 43210, USA howard.1727@osu.edu

While there are promising applications across various domains, it is crucial to acknowledge the potential challenges associated with FER development and integration for robots that utilize the technology to inform their behavior. Recent work in the literature has highlighted concerns related to the manner in which expressive datasets are collected, how annotation protocols are implemented, how performance is evaluated, and how–in turn–certain demographic groups are more susceptible to algorithmic bias [7], [8], [24]. For example, commercial FER systems have been found to exhibit disparate levels of performance based on racial and age identity characteristics in recent years [28], [37]. Additionally, cultural differences in emotional expression and the unique vulnerabilities of certain populations, such as children or the elderly, further complicate the ethical deployment of FER, as false perceptions could lead to inaccurate, harmful or biased robot behavior.[39], [8], [28]. The research community has started to address these concerns, but a fundamental question remains: *Is it ethically justifiable for robots to perceive and interpret human facial expressions at all, and at what cost?*

In [43], Williams raised an ethical question within the HRI community regarding whether robots should possess the ability to represent, recognize, or reason over human identity characteristics and argued that roboticists should not design robots with such capabilities in most cases [43]. These concerns echo related debates over whether AI systems or embodied robots should utilize biometric information, including data related to human faces, which drive use cases such as facial detection and expression recognition [1], [12]. In this paper, we explore potential avenues to address this question and apply existing ethical frameworks to inform developers and promote solutions that empower users to control how their facial expressions are utilized by robots.

In the following sections, we explore four distinct ethical approaches—Ethical Non-Use, Visible Cue Perception, Necessary Informed Consent, and Contextual Appropriateness—that address the complex ethical landscape surrounding FER in HRI. Each approach is critically analyzed in terms of its potential benefits and risks, particularly focusing on invasion of privacy, algorithmic bias, data misuse, and consent mismanagement. The discussion is further enriched by a privacy risk matrix that compares these approaches, highlighting the ethical trade-offs involved. The paper concludes with recommendations for developers and policymakers to ensure that the deployment of FER technology aligns with societal values and ethical standards.

TABLE I

SUMMARY OF ETHICAL APPROACHES TO FER TECHNOLOGY FOR HRI

Approach	Description	Advantages	Limitations	Implementation Difficulty
Ethical Non-Use	Avoids using FER technology altogether.	Eliminates all risks associated with FER; ensures user privacy.	May hinder technological benefits and limit potential applications.	Low – Simple to implement as it avoids using FER technology altogether.
Visible Cue Perception	Allows FER to analyze observable characteristics without inferring emotional states.	Balances interaction quality with privacy; less intrusive.	May still lead to misinterpretation of visible cues; limited by observable data.	Medium – Requires precise interpretation of visible cues without making unfounded inferences.
Necessary Informed Consent	Uses FER technology only with explicit user consent.	Empowers users with control over their data; promotes transparency.	Requires robust consent mechanisms; may be challenging in dynamic environments.	High – Consent processes need to be robust, clear, and adaptable to various contexts.
Contextual Appropriateness	Employs FER based on contextual guidelines and ethical standards.	Adapts FER use to specific contexts, balancing benefits and ethical concerns.	Complex to implement; requires continuous oversight and updates.	High – Requires dynamic adaptation to context, ongoing monitoring, and ethical guidelines.

II. ETHICAL NON-USE

One proposed design pathway, which we refer to as ethical non-use, advocates for the complete avoidance of FER technology in robotics. Proponents of this approach argue that the inherent risks associated with FER—particularly those related to privacy invasion, emotional manipulation, and algorithmic bias—outweigh any potential benefits [3], [4]. The close relationship between a person's emotional state and their facial expressions presents the risk of misinterpretation or unwanted surveillance, potentially leading to breaches of user privacy. Additionally, the capacity for algorithmic harm perpetuated by bias within FER systems is a strong argument for ethical non-use [25], [28].

By avoiding the representation or analysis of facial or identity characteristics entirely, developers could avoid the risk of robots infringing upon users' human rights via FER technology entirely (see Table II). However, this approach may hinder the adoption of social robots, as FER has been shown to play a crucial role in enhancing a robot's perceived emotional intelligence during interaction [38]. Williams further postulates that a robot which is prohibited from recognizing or reasoning over identity characteristics such as race may be accused of racism due to its inherent "colorblindness," which could ignore the social realities of race [43]. In a similar vein, a robot that is restricted from recognizing or interpreting facial expressions might be perceived as lacking essential social skills or emotional intelligence, thereby diminishing its effectiveness in social contexts.

III. VISIBLE CUE PERCEPTION

A more moderate stance is reflected in the Visible Cue Perception approach, where robots can be permitted to perceive and analyze outward and observable characteristics of a user without attempting to infer innate emotional states from these observations. This approach hinges on the ethical distinction between observable characteristics and expressions (e.g., eye color, a smile, or a frown) and the unobservable emotional states that lie beneath them [26], [22], [7]. Since these observable characteristics are publicly displayed, this approach argues that they can be ethically interpreted by robots, minimizing the risk of privacy invasion (see Table II). However, it is crucial that the deployment is implemented with transparency, ensuring that users understand the limits of the robot's FER capabilities, are aware of when they are being observed, and know how their data is being used and stored [11], [31].

The functional benefits of this approach are evident; robots can improve their interaction quality by responding to visible cues, which can be particularly useful in sectors like customer service or education [18], [15]. At the same time, ethical safeguards must be implemented by developers to prevent robots from overstepping their intended scope and inferring unfounded conclusions about a user based solely on visible observations (i.e., inferring the state of *happiness* from detecting that a user is smiling). This approach aims to maintain a clear boundary between what is observable and what might be inferred, ensuring that the interpretation of visible cues does not lead to unwarranted assumptions about a user's internal emotional state or identity.

IV. NECESSARY INFORMED CONSENT

A Necessary Informed Consent ethical approach emphasizes user autonomy by advocating for FER to be employed only with explicit user consent. This aligns with the concept of explicit ethical agency, where individuals actively decide when and how their data, in this case their facial expressions, are used by robots [43]. Central to this approach is the principle that users should have full control over their engagement with FER technology, respecting their right to privacy and their control over their own personal data [21]. In practice, this could involve users opting in to FER functionality during specific interactions or in particular contexts. This framework is an increasingly popular model used for managing AI features for intelligent products and services [23], [2]. For example, a user may choose to allow a robot to analyze their facial expressions during an assistive therapy session but opt-out in other scenarios. By empowering users to make

TABLE II
PRIVACY RISK MATRIX FOR FER TECHNOLOGY IN HRI
Note: Green cells represent low risk. Yellow cells indicate medium risk. Red cells indicate high risk.

Risk	**Ethical Non-Use**	**Visible Cue Perception**	**Necessary Informed Consent**	**Contextual Appropriateness**
Invasion of Privacy	Low – No data is collected or processed, eliminating privacy concerns.	Low – Perception of visible cues reduces the privacy with less sensitive data.	Low – User consent limits data collection to specific, approved contexts.	Medium – Privacy is context-dependent; some situations may involve higher privacy risks.
Algorithmic Bias	Low – No data processing means no opportunity for biased algorithms.	Medium – Bias can still arise in interpreting visible cues, affecting overall outcomes.	Medium – Although consent is obtained, biases in the system can still affect outcomes.	Low – Though difficult, contextual safeguards can be tailored to minimize bias for specific populations.
Data Misuse	Low – No data is available to be misused.	Medium – Visual cues can be used as proxies for complex concepts (e.g., emotional state, mood, personality)	Low – Consent-driven use reduces the likelihood of data being mishandled–with legal implications.	Low – Data misuse risks are often managed by context-specific policies.
Consent Mismanagement	Low – No data collection negates the need for consent.	Medium – Users may not be fully aware of the implications of visible cue analysis, leading to potential consent violations.	Low – Explicit user consent is obtained and respected, reducing risks.	Medium – Implicit consent in varying contexts could lead to misunderstanding or unintended harm.

these decisions themselves, this approach respects individual preferences, promotes transparency in HRI, and encourages increased AI literacy.

However, the success of this approach depends on ensuring that users are fully informed about what FER entails, including its potential risks and benefits [2]. Transparent communication from developers and robotic systems is essential to obtain informed consent. This approach could foster greater trust in robots, as users feel more in control of their interactions. It also allows for more personalized experiences. Nevertheless, the implementation of this approach poses challenges, particularly in dynamic environments where user consent must be continually managed and updated [23]. The risk for algorithmic bias also remains, emphasizing the need for continuous monitoring and regulation to supplement the overall approach, as summarized in Table II.

V. CONTEXTUAL APPROPRIATENESS

Contextual Appropriateness would involve designing FER systems that autonomously integrate ethical considerations based on the specific context in which they operate, eliminating the need for explicit consent or input from the user at each interaction. In this framework, robots function as implicit ethical agents, with FER technology tailored and deployed selectively, guided by predefined ethical guidelines embedded within the system [43]. These guidelines are designed to balance the potential benefits of FER with the need to respect privacy and prevent misuse, tailored to the specific needs of the user population. For instance, in research settings, robots equipped with FER capabilities are often developed to excel within controlled environments and for specific user groups [40], [29]. In these contexts, the robot's decision-making processes are largely pre-configured by contextual factors, ensuring that FER is employed only when necessary and appropriate according to predefined ethical standards. However, extending this behavior beyond controlled environments presents significant challenges and may be an insurmountable task.

This approach alleviates the need for users to make real-time decisions about FER use, thereby reducing their cognitive burden. The system's ethical design ensures that FER is applied in a manner consistent with societal values and legal norms while enabling the robot to perform its functions effectively. Nonetheless, the absence of explicit user consent at each interaction requires rigorous oversight to guarantee that the ethical guidelines governing FER use are robust, transparent, and subject to continuous review and improvement. In practice, such implicit ethical agents remain more of an aspirational goal than a current reality [33], [43].

Tables I and II provide a comprehensive overview of the ethical approaches discussed in this paper. Table I summarizes the four approaches—Ethical Non-Use, Visible Cue Perception, Necessary Informed Consent, and Contextual Appropriateness— while highlighting their key characteristics and their implications for FER technology in human-robot interactions. Each approach is evaluated based on its stance on data usage, consent, and ethical boundaries, providing a clear comparison of their core principles. Table II presents a privacy risk matrix that visually represents the relative privacy risks associated with each ethical approach. The matrix evaluates risks that are directly relevant to FER systems: Invasion of Privacy, Algorithmic Bias, Data Misuse, and Consent Mismanagement. These risks were chosen based on their impact on user rights and data integrity, reflecting concerns highlighted in the literature and discussed within the ethical approaches [21]. We categorize these risks as relatively low, medium, or high, helping to illustrate how each approach addresses risks and the potential trade-offs involved in implementing FER technology for HRI.

VI. Discussion

Williams's cautionary stance on robots' evolving capabilities serves as a reminder of the growing need for ethical oversight, as we have already witnessed instances where technology has gone astray with broader societal implications [43], [5], [36], [9]. Recent literature has highlighted the potential misuse of face perception technology, raising concerns about privacy infringement, intellectual property violations, and civil rights abuses [45], [17], [4], [3]. The domains of policing and surveillance typically face the most scrutiny regarding biased technology, but concerns also arise in areas such as education and healthcare. In these domains, the data-driven systems often reflect the imperfections of our world rather than the fair and equitable society we aspire to achieve [27], [34], [5]. These are all pressing concerns facing the AI field, as premature and under-regulated face-perception technologies may exacerbate systematic inequities and perpetuate social injustices.

A. Progress in FER Regulation & Evaluation

Efforts to address these issues have led to some progress in governance related to AI technologies globally [10], [25]. For example, some current regulations, such as GDPR and the recently passed AI Act in the European Union, impose strict requirements on data privacy and user consent, which directly impact the ethical use of FER [14]. However, the rapid pace of innovation often results in regulation playing catch-up rather than taking proactive measures. Therefore, it is imperative for all stakeholders involved in the design and developmental process to prioritize the ethical considerations of their technologies for all users. Some emerging approaches to address these challenges include actionable auditing and benchmarking [35], [19], [24], improved bias detection and mitigation strategies [42], and proactive consideration of regulations for AI expression recognition systems, which could potentially become more invasive with higher stakes [32]. The intersection of FER with biometric data regulations introduces additional layers of legal scrutiny that social roboticists must also consider.

B. Choosing an Ethical Framework for Integrating FER

In light of the diverse ethical considerations currently surrounding FER technology, it is important for developers, stakeholders, and policymakers to select an approach or combination of approaches that aligns with both the technical capabilities of the robot and the ethical standards of their specific applications. For contexts where user privacy and data security are paramount, the *ethical non-use* approach may simply be the most appropriate, as it eliminates risks associated with data collection and potential misuse. However, if the benefits of FER technology are deemed significant for the application's success, a *necessary informed consent* approach offers a robust framework for ensuring user agency and transparency, provided that consent mechanisms are well-implemented and continually managed. The *visible cue perception* approach can be a viable middle ground, balancing interaction quality with ethical concerns, but robots would need to process such data in a way that does not conflate visible facial cues with more complex concepts like emotional state, mood or personality. Finally, the *contextual appropriateness* approach, while aspirational, presents a forward-looking strategy that integrates ethical guidelines within specific operational contexts, but it requires rigorous oversight and constant evaluation to ensure adherence to ethical standards. Ultimately, the choice of approach should be guided by a careful assessment of the application's goals, the potential risks involved, and the commitment to upholding user rights and fostering user trust. By adopting a thoughtful and contextually aware approach, stakeholders can contribute to the responsible development and deployment of FER technology in HRI.

1) Implementation: Roboticists implementing *ethical non-use* should ensure that FER capabilities are disabled by default. For *visible cue perception*, systems could be designed to process facial cues learned without traditional emotional embeddings. A *necessary informed consent* approach would integrate consent protocols where robots must obtain and routinely evaluate user permission before engaging in FER. *Contextual appropriateness* would involve pre-determined guidelines developed in collaboration with psychologists, legal experts, and ethicists, where robots adapt FER capabilities to the specific context based on these expert-defined criteria, ensuring that the technology aligns with both ethical and legal standards.

C. Future Research Directions

As robots enter more dynamic environments and FER technology continues to evolve, future research should focus on developing and evaluating robust methodologies for implementing and assessing ethical approaches in real-world scenarios. This includes investigating effective methods for robots to determine and integrate user preferences regarding informed consent for FER algorithms and enhancing transparency about user data usage. Additionally, interdisciplinary research combining insights from ethicists, psychologists, and technologists can refine ethical frameworks and improve FER systems' adaptability in diverse HRI contexts. Addressing these areas will promote the responsible deployment of FER technology and help address the challenges highlighted in this paper.

VII. Conclusion

In conclusion, the integration of FER for HRI demands careful ethical consideration. Whether through ethical non-use, visible cue perception, necessary informed consent, or contextual appropriateness, each ethical approach addresses crucial concerns like privacy, bias, and user control. As robots are integrated into various sectors, it is essential to balance the benefits of FER technology with robust ethical guidelines. By fostering transparency, ensuring user autonomy, and implementing thoughtful oversight, we can enhance human-robot interactions while upholding fundamental rights to privacy, equality, and agency.

ACKNOWLEDGMENT

This research is partially supported by funding from the Linda J. & Mark C. Smith Endowed Chair in Bioengineering at Georgia Tech, the NSF GRFP under Grant No. DGE-1650044, the Alfred P. Sloan Foundation MPHD Program under Grant No. G-2019-11435, NSF Award No. 1849101, and the Amazon CoRo Fellowship Program.

REFERENCES

[1] Denise Almeida, Konstantin Shmarko, and Elizabeth Lomas. The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of us, eu, and uk regulatory frameworks. *AI and Ethics*, 2(3):377–387, 2022.

[2] Adam J Andreotta, Nin Kirkham, and Marco Rizzi. Ai, big data, and the future of consent. *Ai & Society*, 37(4):1715–1728, 2022.

[3] Allison Macey Banzon, Jonathan Beever, and Michelle Taub. Facial expression recognition in classrooms: Ethical considerations and proposed guidelines for affect detection in educational settings. *IEEE Transactions on Affective Computing*, 15(1):93–104, 2023.

[4] Lindsey Barrett. Ban facial recognition technologies for children-and for everyone else. *BUJ Sci. & Tech. L.*, 26:223, 2020.

[5] Ruha Benjamin. *Race after technology: Abolitionist tools for the new Jim code*. John Wiley & Sons, 2019.

[6] Carmen Bisogni, Aniello Castiglione, Sanoar Hossain, Fabio Narducci, and Saiyed Umer. Impact of deep learning approaches on facial expression recognition in healthcare industries. *IEEE Transactions on Industrial Informatics*, 18(8):5619–5627, 2022.

[7] De'Aira Bryant, Siqi Deng, Nashlie Sephus, Wei Xia, and Pietro Perona. Multi-dimensional, nuanced and subjective-measuring the perception of facial expressions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20932–20941, 2022.

[8] De'Aira Bryant and Ayanna Howard. A comparative analysis of emotion-detecting ai systems with respect to algorithm performance and dataset diversity. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 377–382, 2019.

[9] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR, 2018.

[10] Elif Kiesow Cortez and Nestor Maslej. Adjudication of artificial intelligence and automated decision-making cases in europe and the usa. *European Journal of Risk Regulation*, 14(3):457–475, 2023.

[11] Claudia Cuador. From street photography to face recognition: Distinguishing between the right to be seen and the right to be recognized. *Nova L. Rev.*, 41:237, 2016.

[12] Laurence Devillers. Human–robot interactions and affective computing: The ethical implications. *Robotics, AI, and Humanity: Science, Ethics, and Policy*, pages 205–211, 2021.

[13] Anukriti Dureha. An accurate algorithm for generating a music playlist based on facial expressions. *International Journal of Computer Applications*, 100(9):33–39, 2014.

[14] Mateja Durovic and Tommaso Corno. The privacy of emotions: From the gdpr to the ai act, an overview of emotional ai regulation and the protection of privacy and personal data. *Privacy, Data Protection and Data-driven Technologies*, pages 368–404, 2025.

[15] Bei Fang, Xian Li, Guangxin Han, and Juhou He. Facial expression recognition in educational research from the perspective of machine learning: A systematic review. *IEEE Access*, 2023.

[16] Chiara Filippini, David Perpetuini, Daniela Cardone, and Arcangelo Merla. Improving human–robot interaction by enhancing nao robot awareness of human facial expression. *Sensors*, 21(19):6438, 2021.

[17] Timnit Gebru. Race and gender. *The Oxford handbook of ethics of AI*, pages 251–269, 2020.

[18] M Rosario González-Rodríguez, M Carmen Díaz-Fernández, and Carmen Pacheco Gómez. Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction through emotions. *Telematics and Informatics*, 51:101404, 2020.

[19] Laura Gustafson, Chloe Rolland, Nikhila Ravi, Quentin Duval, Aaron Adcock, Cheng-Yang Fu, Melissa Hall, and Candace Ross. Facet: Fairness in computer vision evaluation benchmark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20370–20382, 2023.

[20] Md Inzamam Ul Haque. A facial expression recognition application development using deep convolutional neural network for children with autism spectrum disorder to help identify human emotions. 2019.

[21] Javier Hernandez, Josh Lovejoy, Daniel McDuff, Jina Suh, Tim O'Brien, Arathi Sethumadhavan, Gretchen Greene, Rosalind Picard, and Mary Czerwinski. Guidelines for assessing and minimizing risks of emotion recognition applications. In *2021 9th International conference on affective computing and intelligent interaction (ACII)*, pages 1–8. IEEE, 2021.

[22] Rachael E Jack, Oliver GB Garrod, Hui Yu, Roberto Caldara, and Philippe G Schyns. Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109(19):7241–7244, 2012.

[23] Meg Leta Jones, Ellen Kaufman, and Elizabeth Edenberg. Ai and the ethics of automating consent. *IEEE Security & Privacy*, 16(3):64–72, 2018.

[24] Kimmo Karkkainen and Jungseock Joo. Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1548–1558, 2021.

[25] Amelia Katirai. Ethical considerations in emotion recognition. *Mental health*, 4(13):14.

[26] Harmanpreet Kaur, Daniel McDuff, Alex C Williams, Jaime Teevan, and Shamsi T Iqbal. "i didn't know i looked angry": Characterizing observed emotion and reported affect at work. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2022.

[27] Michael Kearns and Aaron Roth. *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press, 2019.

[28] Eugenia Kim, De'Aira Bryant, Deepak Srikanth, and Ayanna Howard. 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*, pages 638–644, 2021.

[29] Konstantinos-Filippos Kollias, Panagiotis Sarigiannidis, Christine K Syriopoulou-Delli, George F Fragulis, et al. Implementation of robots in autism spectrum disorder research: Diagnosis and emotion recognition and expression. In *2023 12th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, pages 1–4. IEEE, 2023.

[30] Alejandro Lopez-Rincon. Emotion recognition using facial expressions in children using the nao robot. In *2019 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, pages 146–153. IEEE, 2019.

[31] Andrew McStay. Emotional ai, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society*, 7(1):2053951720904386, 2020.

[32] Andrew McStay et al. 'this time with feeling?': Assessing eu data governance implications of out of home appraisal based emotional ai. *First Monday*, 24(10), 2019.

[33] James Moor et al. Four kinds of ethical robots. *Philosophy Now*, 72:12–14, 2009.

[34] Safiya Umoja Noble. Algorithms of oppression: How search engines reinforce racism. In *Algorithms of oppression*. New York university press, 2018.

[35] Inioluwa Deborah Raji and Joy Buolamwini. Actionable auditing. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 1 2019.

[36] Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. Saving face: Investigating the ethical concerns of facial recognition auditing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 145–151, 2020.

[37] Lauren Rhue. Racial Influence on Automated Perceptions of Emotions. *SSRN Electronic Journal*, dec 2018.

[38] Ulla A Saari, Antero Tossavainen, Kirsikka Kaipainen, and Saku J Mäkinen. Exploring factors influencing the acceptance of social robots among early adopters and mass market representatives. *Robotics and Autonomous Systems*, 151:104033, 2022.

[39] Klaus R Scherer, Elizabeth Clark-Polner, and Marcello Mortillaro. In the eye of the beholder? universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, 46(6):401–435, 2011.

[40] Fabrizio Schiavo, Lucia Campitiello, Michele Domenico Todino, and Pio Alfredo Di Tore. Educational robots, emotion recognition and asd: New horizon in special education. *Education Sciences*, 14(3):258, 2024.

[41] Huan-Huan Wang and Jing-Wei Gu. The applications of facial expression recognition in human-computer interaction. In *2018 IEEE International Conference on Advanced Manufacturing (ICAM)*, pages 288–291, 2018.

[42] Zeyu Wang, Klint Qinami, Ioannis Christos Karakozis, Kyle Genova, Prem Nair, Kenji Hata, and Olga Russakovsky. Towards fairness in visual recognition: Effective strategies for bias mitigation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8919–8928, 2020.

[43] Tom Williams. The eye of the robot beholder: Ethical risks of representation, recognition, and reasoning over identity characteristics in human-robot interaction. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 1–10, 2023.

[44] Huafei Xiao, Wenbo Li, Guanzhong Zeng, Yingzhang Wu, Jiyong Xue, Juncheng Zhang, Chengmou Li, and Gang Guo. On-road driver emotion recognition using facial expression. *Applied Sciences*, 12(2):807, 2022.

[45] Shikun Zhang, Yuanyuan Feng, and Norman Sadeh. Facial recognition: Understanding privacy concerns and attitudes across increasingly diverse deployment scenarios. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 243–262, 2021.