psychology international

MDPI

Review

Ethical Considerations in Emotion Recognition Research

Darlene Barker [1,2], Mukesh Kumar Reddy Tippireddy [2], Ali Farhan [3] and Bilal Ahmed [4,*]

[1] Research and Development, DB Attic Studios, LLC, Bedford, NH 03110, USA; darlene.barker@dbatticstudios.com or dbarker1@rivier.edu
[2] Department of Computer Science, Rivier University, Nashua, NH 03060, USA
[3] Department of Management Information Systems, Keimyung University, Daegu 42601, Republic of Korea
[4] Department of Biomedical Engineering, College of Engineering, Keimyung University, Daegu 42601, Republic of Korea
* Correspondence: ahmed.bilal.bhimber@gmail.com

Abstract: The deployment of emotion-recognition technologies expands across healthcare education and gaming sectors to improve human–computer interaction. These systems examine facial expressions together with vocal tone and physiological signals, which include pupil size and electroencephalogram (EEG), to detect emotional states and deliver customized responses. The technology provides benefits through accessibility, responsiveness, and adaptability but generates multiple complex ethical issues. The combination of emotional profiling with biased algorithmic interpretations of culturally diverse expressions and affective data collection without meaningful consent presents major ethical concerns. The increased presence of these systems in classrooms, therapy sessions, and personal devices makes the potential for misuse or misinterpretation more critical. The paper integrates findings from literature review and initial emotion-recognition studies to create a conceptual framework that prioritizes data dignity, algorithmic accountability, and user agency and presents a conceptual framework that addresses these risks and includes safeguards for participants' emotional well-being. The framework introduces structural safeguards which include data minimization, adaptive consent mechanisms, and transparent model logic as a more complete solution than privacy or fairness approaches. The authors present functional recommendations that guide developers to create ethically robust systems that match user principles and regulatory requirements. The development of real-time feedback loops for user awareness should be combined with clear disclosures about data use and participatory design practices. The successful oversight of these systems requires interdisciplinary work between researchers, policymakers, designers, and ethicists. The paper provides practical ethical recommendations for developing affective computing systems that advance the field while maintaining responsible deployment and governance in academic research and industry settings. The findings hold particular importance for high-stakes applications including healthcare, education, and workplace monitoring systems that use emotion-recognition technology.

Keywords: emotion recognition; research ethics; data autonomy; algorithmic discrimination; observer bias; dual-use bias; psychological impact; affective computing

check for updates

Academic Editor: Marianna Mazza

Received: 25 March 2025
Revised: 9 May 2025
Accepted: 18 May 2025
Published: 29 May 2025

Citation: Barker, D., Tippireddy, M. K. R., Farhan, A., & Ahmed, B. (2025). Ethical Considerations in Emotion Recognition Research. *Psychology International*, *7*(2), 43. https://doi.org/10.3390/psycholint7020043

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Research in emotion recognition consists of the design of algorithms that can recognize the emotional state of a subject from signals such as facial expressions, voice tone, or even physiological signals like EEG and pupillometry. The purpose is to improve the understanding of human emotions in order to design better systems that may find their application in

almost every area of life: healthcare, education, security, or human computer interaction. For example, Banzon et al. investigated the use of facial expression recognition technologies in classrooms for the purpose of generating adaptive learning environments according to students' emotional states in real time, which would help educators to adapt their teaching tactics accordingly (Banzon et al., 2024). However, they also pointed out the existing ethical issues when using affective computing in educational settings, with emphasis on the issues of privacy, informed consent, and the possibility of improper use of students' emotional data. The ethical approaches to human–computer interaction (HCI) research include participatory and user-centered design, as well as transparency for us as researchers to be able to address ethical risks when working with humans (Lazar et al., 2017).

The implementation of emotion-recognition systems through wearable technology, including smartwatches (e.g., Apple Watch, Fitbit), EEG headsets (e.g., Muse), and biometric wristbands, generates multiple ethical concerns about user consent and data privacy and psychological effects. Using checklists focused on the ethics behind the data collection can help ensure users' autonomy and privacy (Behnke et al., 2024).

Foundational research into affective computing through Rosalind Picard's seminal work (Picard, 1997) and human–computer interaction (HCI) studies by Cowie et al. (2001), which investigated the ethical aspects of emotion identification and response technologies. The initial discussions about these technologies emphasized three main concerns, which included user autonomy, emotional privacy, and potential manipulation. Emotion-recognition technologies have spread across healthcare education and security applications during recent years, which has led to an intensified ethical dialogue that researchers policy-makers and ethicists now focus on Banzon et al. (2024); Behnke et al. (2024); Mattioli and Cabitza (2024).

However, there are numerous ethical issues that emerge when implementing emotion-recognition technologies, chief among them being the issues of bias, equity, and the potential for emotional coercion. Various kinds of bias can affect such systems: algorithmic bias happens when machine-learning models learn the wrong emotion from data and often lead to racial or gender bias as well as neurocognitive bias in emotion recognition Denecke and Gabarron (2024). Data collection bias happens when some groups are underrepresented, which results in the development of models that are not very accurate when applied to a diverse population Kargarandehkordi et al. (2024). Observer bias poses another challenge since cultural features affect the way human annotators assign emotional labels Kargarandehkordi et al. (2024).

The proposed framework includes structural safeguards through adaptive consent mechanisms and real-time user feedback loops to guarantee transparency and user agency and emotional well-being. In addition, the issues of bias in interpretation, as well as the feedback loops, only worsen the ethical problems since they enable the faulty emotion-recognition outputs to affect real-life decisions in education, healthcare or security (Kargarandehkordi et al., 2024).

Psychological and cognitive biases may also have adverse impacts on neurodiverse people or those with cognitive disabilities, which in turn may adversely affect the efficacy of these technologies for these users (Kargarandehkordi et al., 2024). The dual-use bias, however, poses another set of ethical issues since emotion-recognition technologies can be used for a variety of unethical practices, including surveillance, marketing, or social control (Elyoseph et al., 2024).

Another ethical issue is regulatory bias, which leads to unsystematic ethical standards as AI governance may focus on specific sectors or countries and fail to address overall ethical issues (Elyoseph et al., 2024).

The use of algorithmic bias is particularly of concern because emotion-recognition models are poorly prepared to work with cultural and demographic diversity and tend to produce inaccurate classifications that are discriminatory in nature. A review of transformer-based multimodal emotion-recognition architectures also showed that varied datasets improve the accuracy as well as the fairness of the models (Pavarini et al., 2021). Similarly, a different work showed the necessity of integrating emotional intelligence into AI systems, but also highlighted possible psychological risks of constant emotional monitoring, such as loss of autonomy and exploitation (Kissman & Maurer, 2002). The authors recommend robust regulatory frameworks, improved consent mechanisms, and effective bias mitigation strategies (Mattioli & Cabitza, 2024). Additionally, since the HCI research presents the best practices of complying with the guidelines of ethical considerations in the implementation of HCI systems, the user's autonomy, informed consent, and transparency of the system design should also be maintained (Lal & Neduncheliyan, 2024).

Emotion-recognition technologies are expanding rapidly across many fields including healthcare, marketing, education, and security, which present many opportunities to improve the user experience and the relation between human and computers. These systems promise transformative benefits in customer service automation, mental health diagnostics, and security surveillance, among high profile cases (Dominguez-Catena et al., 2024; Sidana et al., 2024). However, all these advancements raise significant ethical concerns, especially for privacy, bias, and emotional manipulation. These issues are compounded by the sensitivity of the biometric data and the need for comprehensive ethical guidelines in emotion-recognition research, given recent cases of data misuse and privacy breaches, as well algorithmic biases (IEEE, 2024; Lakkaraju et al., 2024). The privacy concerns are further exacerbated by the fact that biometric data is often collected, stored, and used without an individual's knowledge (Dominguez-Catena et al., 2024; Durachman et al., 2024). In addition, current emotion-recognition algorithms have problems with cultural and demographic shifts, which can result in misclassification (Dominguez-Catena et al., 2024; Kotian et al., 2024; Sidana et al., 2024).

These essential ethical issues are explored in this paper, which offers a systematic framework for the responsible application of emotion-recognition technologies. We review privacy risks, algorithmic fairness, and the psychological effects of being monitored as a person. We suggest transparency, informed consent, fairness, and regulatory oversight to protect individual rights while, at the same time, optimizing societal value (Durachman et al., 2024; Singh & Kamboj, 2024).

1.1. Problem Statement

We are referring to technologies that collect data via devices including sensors or cameras and statistical models to identify and/or decode the emotional state of a person. This can be achieved in different ways, such as through electroencephalogram (EEG), which interprets brain waves, pupillometry, which is the study of the pupil response and gaze behavior, facial expression coding, and other physiological measures (Barker & Levkowitz, 2024b). These technologies have great potential in various applications including education, healthcare, and human computer interaction to design emotionally aware systems to enhance learning, improve healthcare, and design better user interfaces (Banzon et al., 2024; Florestiyanto et al., 2024; Schrank et al., 2014). Nevertheless, the growing use of these tools brings about ethical issues such as privacy, informed consent, bias of algorithms, and cultural sensitivity (Behnke et al., 2024; Kissman & Maurer, 2002; Vlek et al., 2012).

Hitherto, the development of emotion-recognition technologies has been widely discussed; nevertheless, the social and ethical implications of such technologies have not been fully addressed. Important issues include protecting the privacy of biometric data col-

lected and stored, explaining the processes that require consent and obtaining meaningful consent, and dealing with the biases that affect marginalized populations (Behnke et al., 2024; Hazmoune & Bougamouza, 2024; Kissman & Maurer, 2002). These challenges shall not be met by simply designing ethical tools but by incorporating ethical principles into the design of the systems themselves in order to tackle the perceived issues and increase acceptance (Muir et al., 2019; Ng, 1997).

1.2. Motivation

The technology for emotion recognition shows great potential to transform educational settings and healthcare facilities and human–computer interaction systems. In education, they make it possible to develop smart education environments that adapt to students' emotions, thus improving the personalized learning processes (Banzon et al., 2024; Jagtap et al., 2023). In healthcare, these technologies can substantially enhance the diagnosis and treatment of mental health disorders through the real-time analysis of patients' emotional states, which can lead to the development of specific treatment plans based on affective changes (Lal & Neduncheliyan, 2024; Schrank et al., 2014; Vlek et al., 2012). These technologies can lead to the development of empathetic and intuitive human–computer interfaces, which can increase user engagement and satisfaction (Hazmoune & Bougamouza, 2024; Obaigbena et al., 2024).

The problems of collecting and storing biometric data and its further use are still relevant, which raises questions among participants concerning the use of the data and its possible misuse (Behnke et al., 2024; Kissman & Maurer, 2002; Mattioli & Cabitza, 2024). The integration of emotion recognition into surveillance and marketing strategies increases the likelihood of its misuse. Deployment is further hampered by issues of informed consent, especially with passive emotion tracking technologies (Rizzo et al., 2003; Smith et al., 2025). Also, algorithmic biases sustain oppression of marginalized groups through the representation of minority groups in training datasets (Dominguez-Catena et al., 2024; Muir et al., 2019; Sidana et al., 2024). In addition, the cultural variability of emotional expressions and ethical requirements call for region-specific ethical guidelines to avoid the unfair and culturally insensitive treatment of participants (Han et al., 2021; Ng, 1997).

This paper is prompted by the necessity to mitigate these ethical issues with a clear framework for the development and deployment of responsible emotion-recognition technology (Pavarini et al., 2021; Rizzo et al., 2003). Lazar argued that "participants must be treated fairly and with respect" as a principle which is informed by this research to protect the user's autonomy and privacy in every emotion-recognition application (Lazar et al., 2017).

2. Related Works

Emotion-recognition technologies are gradually finding their application in almost every sphere of life including educational institutions and mental healthcare (Banzon et al., 2024; Florestiyanto et al., 2024). A vast review of the literature on automatic facial emotion recognition (FER) conducted for this article pointed to the problems of privacy invasion and equity in real-life scenarios (Mattioli & Cabitza, 2024) while in Behnke et al. (2024), the authors suggested an ethical framework for affective computing with wearables, focusing on issues of the subject's control and data protection. In the educational environment, another study investigated the implications of implementing facial expression recognition in schools, claiming that the process of monitoring students' emotions poses ethical concerns with regard to consent and the possible misuse of personal emotional data (Banzon et al., 2024). Together, these works underscore the importance of ethical frameworks for the use of emotion-recognition technologies (Banzon et al., 2024; Behnke et al., 2024; Mattioli & Cabitza, 2024).

In their paper, Mattioli and Cabitza (2024) offer a clear and concise categorization of the major ethical issues associated with automatic facial emotion recognition, which include privacy, algorithmic bias, psychological impact, and transparency. Research is still ongoing, e.g., the creation of a standardized multimodal database like Harmonia (Anthireddygari & Barker, 2024), and emotion-recognition system, such as Thelxinoë (Barker & Levkowitz, 2024a, 2024b), which highlight changing ethical concerns regarding privacy, informed consent, and algorithmic fairness in emotion-recognition research.

One of the most significant problems in emotion-recognition research is the problem of algorithmic bias, specifically in facial expression recognition. A study was conducted to determine the existence of racial bias in FER models, which showed that the accuracy of emotion classification differs greatly among different demographic groups (Sidana et al., 2024). However, another study further explored demographic bias in facial expression recognition datasets and presented fairness-aware data augmentation strategies to improve the bias in model predictions (Dominguez-Catena et al., 2024). In the last study, we examined cross-cultural biases in emotion recognition and found that cultural differences affect the way that people display and convey emotions (Han et al., 2021). These studies also point to the problem of the lack of diversity in the datasets used for training the models, which results in the development of the systems that are supposed to be used by a diverse population. So, we also have to protect the input to make sure that we are not using poor data, and we have to include more data from a diverse population.

Emotion-recognition technologies are great, but there are some issues that are not connected with the technologies themselves, for example, psychological and social effects. The application of emotion-monitoring technologies in psychotherapy raises issues on patient's rights, such as autonomy and confidentiality (Muir et al., 2019). In the therapeutic setting, another study highlighted some potential risks of using memory control strategies in emotion regulation, and the potential negative impact on individuals' well-being (Elsey & Kindt, 2016). Another study investigated the role of spirit in mental health and substance use disorder treatment, with a particular focus on the ethical issues involved in incorporating emotion-detection devices into psychological health care (Kissman & Maurer, 2002). These studies show the wider consequences of having emotional surveillance carried out to the level that can be used to control how individuals feel and eventually make them do things they may not want to do.

To deal with these ethical issues, researchers have stressed the importance of ethical data collection and reporting practices in research (Lazar et al., 2017). A second work examined the ethical implications of AI-based emotion recognition, calling on policymakers to develop regulatory measures to safeguard users' rights (Knight, 2024). Taken together, these studies highlight the importance of upholding ethical standards in emotion-recognition research.

Banzon suggested ethical recommendations for the school setting, with focus on the issues of privacy and informed consent (Banzon et al., 2024). Behnke offered an approach to the problem of privacy, data management, and user engagement in wearable emotion-recognition devices (Behnke et al., 2024). However, Rizzo et al. (2003) pointed out that there is a need to strengthen the protection of privacy in psychological interventions, and Kaiser et al. (2020) pointed out that the technology can be misused for surveillance and marketing purposes.

Sidana et al. (2024) explained how data-fusion strategies can help reduce bias in facial expression analysis, and Dominguez-Catena et al. (2024) described tools for assessing the demographic bias of a dataset. Emotion recognition without addressing these biases would only serve to perpetuate injustice (Muir et al., 2019; Neethirajan, 2024).

From the literature review, Behnke et al. (2024) stressed the need for well-defined protocols of sharing data to avoid their misuse, while Elsey and Kindt (2016) discussed

the ethical risks of using emotion-manipulation techniques. Also, Ng (1997) discussed the cultural specificity of emotion recognition and suggested that since emotion can be interpreted differently across regions, there is a need for region-specific ethical guidelines to avoid prejudice and discrimination.

The continuous tracking of emotions may result in the loss of autonomy, the development of stigma, or emotional intimidation (Cartwright et al., 2024). Schrank et al. (2014) brought positive psychology into the discussion of ethical emotion analysis, while Pavarini et al. (2021) investigated the use of participatory arts for ethical mental health interventions.

Previous research has investigated specific aspects of emotion recognition, including studies using wearable devices (Behnke et al., 2024), emotion recognition via AI (Jindal & Singh, 2025), and implications for mental health (Pavarini et al., 2021; Schrank et al., 2014). However, these studies have identified biases in such systems (Dominguez-Catena et al., 2024; Vlek et al., 2012), thus the need for more robust frameworks for fairness.

In their paper, Mattioli and Cabitza (2024) offer a clear and concise categorization of the major ethical issues associated with automatic facial emotion recognition, which include privacy, algorithmic bias, psychological impact, and transparency.

This conceptual work fills these gaps by reviewing the literature and recommending guidelines for ethical design and implementation. Through a systematic review of the literature on privacy issues, informed consent, and cultural aspects of ethical perceptions, we seek to contribute to the development of future studies and support the formulation of integrated ethical frameworks (Pavarini et al., 2021; Rizzo et al., 2003).

3. Conceptual Framework

Emotion-recognition systems have recently attracted much attention but also have the potential to apply to various sectors such as education, healthcare, and security. Nevertheless, these developments also imply critical ethical issues that demand clear guidelines for their use. The research presents a conceptual framework which unites technical aspects with ethical elements and regulatory requirements to support fair and transparent responsible emotion-recognition technology development.

The emotion-recognition system workflow diagram presented in Figure 1 shows the main stages, which start with data collection and end with classification and results reporting. The different phases present unique ethical challenges which require specific design choices for their resolution. The ethical collection of data requires researchers to protect participant autonomy while using clear participation agreements and implementing sufficient data-minimization methods. Developers need to actively watch out for both representational distortions and demographic performance gaps during feature extraction and model-training processes (Benouis et al., 2024; Dominguez-Catena et al., 2024; Mattioli & Cabitza, 2024). System deployment requires the implementation of fairness checks and context-aware design and clear communication protocols to maintain both trust and integrity.

Figure 2 graphically represents the ten key ethical issues discussed in this paper as the basic principles of ethical emotion-recognition research. At its heart is an ethical framework that highlights the need for an integrated approach to address the issues inherent in developing technologies that embed themselves in human lives. With a focus on the participants, the surrounding components, such as informed consent, participant control, and psychological impact, address the rights and well-being of research participants. Technical and societal elements, for instance, bias and fairness, privacy and data protection, and transparency and contextual awareness are critical. The outer nodes also include commercialization, regulatory oversight, and dual-use concerns that capture the broader institutional and systemic factors. The layout of the two sections together shows that the

ethical practice in this domain is multi-dimensional and therefore needs coordinated attention to the development, deployment, and governance of emotion-recognition technologies to be useful and safe for the public.



Figure 1. This figure illustrates a typical emotion-recognition system workflow which shows essential decision points from data collection to processing and output stages while requiring ethical safeguards such as participant autonomy and system transparency and accountability (Barker & Levkowitz, 2024a).

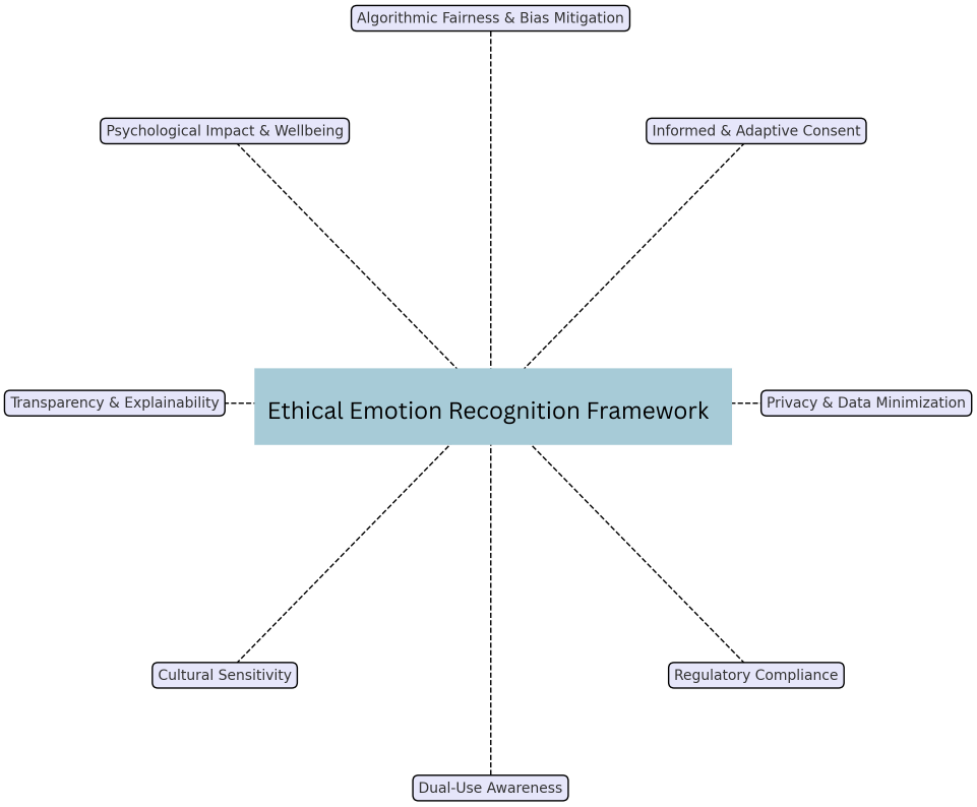


Figure 2. Conceptual Framework for ethical emotion recognition. The figure illustrates the ethical dimensions of privacy, consent, fairness and cultural sensitivity in relation to the central framework established in this study. The links in this figure represent methodological principles that guide the ethical development and implementation of emotion-recognition technologies in different application domains.

In order to solve the ethical issues mentioned in the literature on emotion recognition systems, this paper proposes a conceptual framework that combines technical, ethical, and regulatory aspects. With the help of the recent studies on algorithmic bias, privacy, and the psychological effects of emotion-recognition technologies, the framework is designed to foster the sustainable design and use of these systems. While developing this framework of the ethical issues in research in emotion recognition, the following areas need to be addressed: informed consent (Di Dario et al., 2024; Han et al., 2021), privacy and data protection (Benouis et al., 2024; Jindal & Singh, 2025; Khan et al., 2024; Lin & Li, 2023), bias and fairness (Dominguez-Catena et al., 2024; Mattioli & Cabitza, 2024; Mehrabi et al., 2021; Parraga et al., 2025), psychological impact on participants (Han et al., 2021), dual-use concerns and misuse (Han et al., 2021), transparency and accountability, cultural and

contextual sensitivity (Mattioli & Cabitza, 2024), commercialization and profit making, participant control over the data (Behnke et al., 2024), and regulatory compliance and institutional oversight (Han et al., 2021; Mattioli & Cabitza, 2024). All of these areas are equally crucial in making sure that the participants are well informed and protected in an emotion-recognition study because we are dealing with emotions, some of which are intentional and others not, and some of which are controllable while others are not.

3.1. Informed Consent

According to Behnke, researchers must explain the entire data collection process including data usage, storage practices, and sharing procedures while participants maintain the right to withdraw their participation at any moment without facing any negative consequences (Behnke et al., 2024). Research studies that lack proper participant agency mechanisms risk violating individual autonomy. The complex nature of affective computing technology makes it difficult for many participants to understand the complete meaning of their involvement (Banzon et al., 2024). Di Dario stresses the need for educational institutions to develop ethical guidelines that start by explaining emotional data practices to participants while maintaining open communication about their rights and potential risks (Di Dario et al., 2024).

Smith et al. (2025) emphasize the necessity of user-driven consent management in decentralized virtual environments, and the need for systems that prioritize autonomy and clear communication. Their work requires adaptive frameworks that allow individuals to control their emotional data access, interpretation, and retention at all times. The always-on nature of emotion recognition in immersive platforms like VR means that initial agreements are not sufficient; instead, consent must evolve into an interactive, context-sensitive process. Protecting affective privacy in these settings requires tools that support real-time user control alongside transparent, user-friendly data-governance mechanisms.

Burgio's research shows that people with mild cognitive impairment may have trouble reading facial expressions and this can also be a problem with fully understanding data-related participation agreements. Neurocognitive limitations, as noted in their findings, may make it difficult for a participant to understand how their emotional data will be collected, stored, and used. These findings suggest the need for simplified, accessible documentation and supportive communication strategies to ensure ethical inclusion and informed participation in emotion-recognition studies (Burgio et al., 2024).

In AI-based emotion-recognition systems used for mental health interventions, Denecke and Gabarron stress on the importance of understandable consent procedures (Denecke & Gabarron, 2024). The study reveals that users may not understand all the ways in which their sentiment data is being evaluated, which raises concerns about user autonomy and decision making.

In AI-based recruitment platforms, informed consent has to involve traditional data protection issues as well as the consequences of the emotion analysis of the video interviews. Jogi points out that job applicants may not understand how their microexpressions are being captured and analyzed to determine their authenticity, which requires explicit disclosure and participant control over their data (Jogi et al., 2024).

As a rule, in any research, the participants should be aware of and willing to participate in the study. This is especially important for emotion-recognition studies, which imply the collection of biometric and behavioral data. Previous studies have also highlighted the importance of simple, easily understandable, and legally valid mechanisms for obtaining informed consent (Di Dario et al., 2024).

Participants must be informed about:

1. What data is being collected?

2. How their data will be used.
3. If they have the right to withdraw at any time.

The research highlights that emotion-recognition studies become dangerous to individual autonomy and emotional data privacy when proper participant protection measures are not established. Researchers must provide detailed instructions about the permanent management of emotional data after the initial consent process. The researchers should explain their data storage periods and their methods for secure data deletion and the circumstances that allow data re-use. These practices strengthen participants' trust and align with evolving privacy regulations governing biometric and affective data (Denecke & Gabarron, 2024).

3.2. Privacy and Data Protection

Emotion-recognition systems depend on the use of very intimate information such as facial expressions, voice frequencies, EEG signals, and autonomic responses. If this data is not properly disposed of, it can cause a great deal of harm.

Benouis et al. (2024) show a real-world application of privacy-by-design in emotion-recognition systems. The framework employs federated learning and differential privacy to store emotional data on user devices while distributing only anonymized aggregated model updates. The decentralized method reduces the exposure of sensitive affective information to a great extent and demonstrates the requirement of protective mechanisms in system architecture from the start.

Fry conducts a legal review of emotion data processing risks while providing GDPR and other data-protection framework compliance recommendations (Scannell et al., 2024). The protection methods include anonymization methods that stop personal identification along with secure storage methods to reduce leakage and role-based access control. Systems can lack proper protection mechanisms, which makes them vulnerable to various serious threats including data breaches and identity theft as well as unauthorized emotional surveillance (Benouis et al., 2024; Jindal & Singh, 2025; Khan et al., 2024; Lin & Li, 2023).

The protection of affective data should be a fundamental principle in ethical emotion-recognition research. Emotional signals are inherently sensitive and should be processed with a high level of care to prevent misuse. Khan et al. (2024) and Benouis et al. (2024) highlight differential privacy and federated learning as essential methods to preserve emotional data integrity while maintaining system utility. The European Union's AI Act together with its proposed regulations serves as a basis for legal accountability in system operations (Scannell et al., 2024). Surveillance-driven misuse with far-reaching ethical consequences could occur if protective measures are not implemented (Banzon et al., 2024).

Denecke and Gabarron (2024) highlight that emotional profiling in AI-driven chatbots for mental health becomes problematic when affective cues extend beyond therapeutic applications, thus requiring strict control over emotional information access and application.

New privacy risks are raised by multimodal emotion and engagement recognition systems due to the deployment of highly intimate biometric information. According to Sukumaran and Manoharan, it is crucial to guarantee the protection of data with regards to the capturing of students' affective states from the facial expression, gaze tracking, and eye blinking techniques in online classes (Sukumaran & Manoharan, 2024).

This paper aims to investigate the compatibility of AI in recruitment, and in doing so, it reveals new issues regarding privacy and data protection. Jogi explains how to integrate NLP and emotion analysis to identify microexpressions from video interviews, posing questions on the reliability of biometric data, the possibilities of its misuse through emotional profiling, and legal compliance (Jogi et al., 2024).

Studies involving individuals with cognitive impairments must guarantee very strict data privacy measures. Burgio's research points out the dangers of neuroimaging and affective data collection and calls for stronger privacy measures to prevent the abuse of biometric and neurological data of vulnerable participants (Burgio et al., 2024).

Personalized affective models enhance the accuracy of the model but at the same time, pose new privacy risks. Kargarandehkordi argues that such models entail the storage of the user's affective data, which if not properly protected, can be misused or accessed inappropriately (Kargarandehkordi et al., 2024).

3.3. Algorithmic Bias and Fairness

Emotion recognition faces persistent challenges because algorithms demonstrate inconsistent performance when analyzing different demographic groups. Multiple studies demonstrate that emotion-recognition systems incorrectly identify emotional expressions in people with darker skin tones and men and people from different cultural backgrounds (Dominguez-Catena et al., 2024; Mattioli & Cabitza, 2024; Mehrabi et al., 2021; Parraga et al., 2025). The source of these disparities includes training data that lacks diversity together with the incorrect belief that emotional expressions are universal and insufficient consideration of cultural differences. Researchers need to focus on building inclusive datasets while developing mechanisms to detect representational imbalances in model development and conducting regular fairness assessments during deployment. Through these initiatives, emotion-recognition systems will progress toward reducing harm and enabling fairer applications.

The fundamental problems of algorithmic justice and representation continue to be essential concerns for the field. Facial emotion classifiers demonstrate societal prejudices through their results, which produce racial and gender and ethnic biases according to multiple studies (Dominguez-Catena et al., 2024; Mehrabi et al., 2021; Parraga et al., 2025). Mattioli and Cabitza challenge the "universality hypothesis" because it establishes rigid emotional-expression relationships which hides how emotions manifest differently between cultures (Mattioli & Cabitza, 2024). Parraga demonstrates that deep-learning models trained with unrepresentative populations tend to fail at generalization, which leads to persistent performance gaps between demographic groups (Parraga et al., 2025). The field needs to implement broader data strategies combined with context-aware models and sustained predictive equity audits to address these disparities.

In emotion recognition, algorithmic fairness has to take into account differences in the level of cognitive function. Burgio reported that people with mild cognitive impairment have distinct emotional processing, which means that AI models should be trained on numerous people, including those with neurocognitive disorders, to prevent bias in the classification process (Burgio et al., 2024).

The evaluation of deployed systems for bias in real-world settings requires regular independent algorithmic audits in addition to designing representative datasets. The audits should include fairness metrics across demographic groups and be conducted by third-party evaluators to enhance transparency and accountability (Dominguez-Catena et al., 2024; Mattioli & Cabitza, 2024).

Large-scale emotion-recognition models are usually not accurate in identifying emotions in various populations because of the existing bias in the training data. Kargarandehkordi shows that personalizing the ML models can help reduce such biases by customization of the recognition process to the specific emotional responses of the subject which in turn increase the fairness and decrease the chances of misclassification (Kargarandehkordi et al., 2024).

This paper finds that algorithmic bias in sentiment and emotion analysis can have crucial implications for fairness and equity in mental health care. According to Denecke and Gabarron, it is possible that the use of inaccurate datasets during the training of chatbots can lead to incorrect emotional perception and, therefore, the incorrect categorization of individuals based on their cultural and psychological differences (Denecke & Gabarron, 2024).

Emotion recognition and engagement models are known to be biased towards the data used in their training, which can lead to inaccurate predictions for a given population. In addition, Sukumaran and Manoharan discuss the problem of insufficient data for engagement recognition, with a specific focus on online education, and highlight the importance of expanding the dataset to prevent the exclusion of certain populations in the assessment process with AI-based tools (Sukumaran & Manoharan, 2024).

Automated hiring processes are convenient; however, they may continue existing biases in the training data. Jogi pointed out that ML-based job recruitment and emotion analysis models may prefer some candidates over others and recommended bias auditing and fairness-oriented model reconfiguration in the case of automated hiring (Jogi et al., 2024).

Societal biases are incorporated into and, in some cases, perpetuated by AI systems. Elyoseph argue that GenAI is also biased in learning from human cultures and that there is a need to have a constant watch on the use of AI in emotion recognition (Elyoseph et al., 2024).

3.4. Psychological Impact on Participants

Emotion-recognition research is often conducted with no regard to the psychological consequences of the studies. It is possible that participants will feel uncomfortable with being watched, the system may misidentify their emotions as negative, and they will be unsure what is being done with their feelings. Narimisaei's research looks at the effects of using AI-driven emotion-recognition systems and reveals possible negative impacts on users' mental health (Narimisaei et al., 2024).

Ethical research has to include a de-identification of participants, for example through a debriefing, to make them aware of the impact of the study, to provide information on where participants can find support for their mental health in case they need it, and to ensure that participants are informed of the purposes for which their data will be used. The consequences of not considering the psychological implications of these systems could be adverse, especially in critical settings such as mental health diagnosis.

It is crucial to assess the psychological implications of using the emotion-recognition systems that we design and the data collection and testing processes as part of the user evaluation process (Han et al., 2021; Narimisaei et al., 2024). The psychological implications of emotion-recognition technologies are another ethical concern. Narimisaei et al. (2024) raises concerns about the emotional discomfort that participants may experience whenever their emotions are misunderstood by the AI models. In high stake settings such as education where emotion recognition is applied for student's engagement monitoring, there is a high probability of emotional manipulation and adverse psychological effects (Banzon et al., 2024). According to Muir, it is crucial to incorporate mental health aspects into the design of affective computing systems in order to avoid humiliating and potentially harmful scrutiny of users' emotions (Muir et al., 2019).

People with cognitive impairments are likely to have problems with emotion recognition that may cause them psychological distress when using affective computing systems (Morellini, 2022). These factors should be taken into consideration by the researchers and thus participants should be properly briefed and given mental health information (Burgio et al., 2024).

Autonomous engagement recognition systems are negative in effect to the students' psychological development through constant affective state tracking. Sukumaran and Manoharan explained that students may experience discomfort with the knowledge that their face and eyes are being watched and controlled, and thus, ethical issues in AI-based learning environments (Sukumaran & Manoharan, 2024).

Emotion-recognition models, including those that use a personalized approach, shape users' behavior and perception of the environment. Kargarandehkordi argued that the integration of emotion recognition in the user interface may pose ethical problems as users may rely on the emotional state as identified by AI and this may have adverse psychological implications (Kargarandehkordi et al., 2024).

Elyoseph opined that GenAI offers a way of engaging in emotional interactions in a human-like manner, which in turn results in the user developing an emotional dependence that may have adverse psychological implications (Elyoseph et al., 2024). In emotion-recognition research, this raises ethical questions about the kind of interaction that participants have with the AI models in the study, so that participation is always a positive experience and not a form of exploitation.

The following issues were identified: stress and dependence on the emotion-recognition systems based on AI are becoming a problem. In Denecke and Gabarron (2024), they discussed how users may rely more on chatbots for emotional support and may avoid seeking help from human mental health professionals, which leads to increased social isolation. There is a need to examine the psychological implications of using the emotion-recognition systems that we are developing as well as the data collection and testing processes (Narimisaei et al., 2024).

The effectiveness of emotion-recognition technologies presents another ethical issue. Narimisaei also pointed out that participants may feel angry or upset when their emotions are misdiagnosed or misunderstood by the AI models (Narimisaei et al., 2024). In education, where emotion recognition is used for student engagement analysis, there are concerns of emotional manipulation and other psychological risks (Banzon et al., 2024). According to Muir, it is crucial to incorporate mental health factors into the implementation of affective computing systems to avoid putting users in a position of emotional surveillance or distress (Muir et al., 2019).

People with cognitive impairments who have mild cognitive impairment (MCI) face difficulties in emotion recognition, which results in psychological distress and miscommunication when using affective computing systems (Morellini, 2022). These factors should be taken into consideration by the researchers and therefore participants should be properly briefed and given mental health information (Burgio et al., 2024).

The automated engagement recognition systems are detrimental to the psychological development of students as they monitor and interpret students' emotions. Sukumaran and Manoharan stated that students may experience discomfort with the knowledge that their facial expressions and eye movements are being tracked, and this calls for ethical consideration of the use of AI in learning environments (Sukumaran & Manoharan, 2024).

The interaction between users and emotion-recognition models, especially those that are personalized, affects the two parties in the desired manner. Kargarandehkordi et al. (2024) argued that emotion recognition may pose ethical issues if the user relies on the emotional state provided by the AI, which may have adverse psychological implications.

In emotion-recognition research, this raises ethical concerns about the kind of interaction that participants have with the AI models in the study, in that the interaction should be positive and not negative in any way.

Studies have revealed that people with cognitive impairments have problems in emotion recognition, which may lead to psychological problems when using the affective

computing systems. Burgio also pointed out that these factors should be taken into consideration by the researchers and therefore participants should be properly briefed and given mental health information (Burgio et al., 2024).

The ethical strength of these technologies increases when participants receive psychological support resources for any unexpected emotional distress they experience during or after studies (Pavarini et al., 2021; Schrank et al., 2014). The psychological implications of the use of AI in recruitment are considerable. Jogi discuss issues on stress levels of candidates who are subjected to real time emotion analysis during job interviews and raise ethical concerns about transparency and the mental health of applicants (Jogi et al., 2024).

3.5. Dual-Use Concerns and Misuse

This paper considers emotion-recognition technology as having a dual-use potential, that is, it can be used for better or worse purposes. Although it can benefit mental health care, it can also be used for surveillance (tracking emotions in public spaces), workplace monitoring (monitoring the emotions of employees without their knowledge), and manipulative marketing (understanding emotions to show particular advertisements). This technology needs to be regulated to avoid its misuse (Han et al., 2021). As with the guidelines for ethical deployment of virtual reality suggested earlier, our approach focuses on ensuring openness, voluntary participation, and the involvement of affected parties to avoid the abuse of the technology (Raja & Al-Baghli, 2025).

The dual-use issues and possible misuse of emotion-recognition systems only increase the ethical issues. Although these technologies can be used appropriately in mental health care and human computer interaction, they can also be used for surveillance and behavioral control (Han et al., 2021). Kleinrichert's application of emotion recognition in the workplace for performance management has important ethical implications for employee control and choice (Kleinrichert, 2024). The main focus of the field of affective computing is mainly commercial and this is usually in a bid to achieve profit (Knight, 2024).

Emotion-recognition technology is beneficial in the clinical environment; however, it has its dangers. Narimisaei et al. (2024) explain how the use of sentiment analysis in chatbots can be misused for marketing purposes or surveillance without the user's knowledge, and therefore needs legal framework.

The use of emotion recognition in recruitment platforms also comes with certain risks of dual use. Jogi points out that if emotion-analysis technology is adopted for hiring purposes, it may be used for other purposes such as constant emotional monitoring of employees or selfish workplace analytics (Jogi et al., 2024).

The use of AI in emotion recognition is increasing and this brings with it several problems of abuse. Elyoseph explained that in GenAI, the copying of human behaviors leads to ethical issues in areas such as surveillance, entertainment, or business where the AI-generated emotions are used manipulatively (Elyoseph et al., 2024).

There are dual-use risks in the deployment of AI for engagement tracking. Although these models are useful in helping the teachers know the level of attention of the students, Sukumaran and Manoharan explain that there are ethical issues that may arise when the systems are used for surveillance or punishment, thus raising issues on student's rights and permission (Sukumaran & Manoharan, 2024).

3.6. Transparency and Accountability

In order for emotion recognition to be trusted, it has to be explainable. However, many of the current systems are black box approaches and it is tough to understand how they understand emotions, what data they use, and whether they are biased. To this end,

researchers should provide comprehensive information about their models, use explainable AI (XAI), and make it possible for other people to review the models.

This paper aims to contribute to the discussion of how to achieve transparency and accountability in the development and usage of emotion-recognition technologies. At present, many AI models used for emotion recognition are black box models, which make it hard to understand the reasons behind the classifications made (Hazmoune & Bougamouza, 2024). Cartwright's research also supports the use of XAI techniques to increase the understandability of affective computing systems (Cartwright et al., 2024), and furthermore, external audits and ethical review boards should be established to watch over the proper use of these technologies (Behnke et al., 2024).

The complexity of modern AI emotion-recognition models poses a threat to classic ethical concepts such as accountability and understanding. Elyoseph points out that with the growing role of AI in decision-making processes that were previously the domain of human cognition, it is crucial to address the issue of accountability in the ethical use of such technologies (Elyoseph et al., 2024; Scannell et al., 2024).

The use of interpretable models in AI-based emotion recognition is vital for the ethical use of the technology. Kargarandehkordi recommends classical ML methods over deep learning to enhance the explainability of the model and further stresses the need to make AI decisions more understandable and justifiable (Kargarandehkordi et al., 2024).

This paper aims at contributing to the discussion on how to achieve transparency in emotion-recognition models to increase trust and better implement ethical AI. Sukumaran and Manoharan propose an 'Engagement Indicator' algorithm that combines different biometric features and, therefore, highlight the importance of understanding the method used to calculate the engagement score to avoid covert decisions (Sukumaran & Manoharan, 2024).

3.7. Cultural and Contextual Sensitivity

Emotional expressions are very culture specific. What is the anger face in one culture may mean confusion in another culture (Mattioli & Cabitza, 2024). Most current emotion-recognition systems do not consider this, which results in incorrect emotion-recognition of people from different cultural groups. To increase the cultural relevance of the research, it is crucial that diverse populations are engaged in the data collection process, and the models should be able to interpret the input in the context of the culture (Banzon et al., 2024; Mattioli & Cabitza, 2024).

Cultural and contextual sensitivity is another important aspect of ethical emotion recognition. Facial expressions and emotional gestures are very much a function of cultural norms and practices, yet most current systems are ignorant to this fact (Mattioli & Cabitza, 2024). Han also explored how regulatory focus and emotion-recognition biases differ across cultures, affecting negotiation behaviors and interpersonal interactions (Han et al., 2021). To enhance cultural sensitivity, researchers should design self-adapting models that include the cultural dimensions of emotional display (Banzon et al., 2024).

3.8. Commercialization and Profit Driven Motives

A large number of companies are trying to create emotion-recognition technology for the purpose of commerce and there is no regard to ethics. Ethical risks are linked to the lack of control of users over their data, the surveillance risks of consumer applications, and the use of personal emotions in advertising. It is important that in this rapidly developing area, ethics must be considered alongside commercial interests.

3.9. Participant Control over the Data

Participants should have the right to control their data, i.e., the right to decide who can use the data, how the data will be stored, and whether the data can be deleted (Behnke

et al., 2024). Ethical research should guarantee that participants can opt out at any time, that the data is not shared without the participant's consent, and that the participant knows how the data collected is being used.

3.10. Compliance with the Regulatory Authorities and Institutional Oversight

This paper argues that governments and regulatory bodies should issue specific guidelines on how to design emotion-recognition systems responsibly (Han et al., 2021; Mattioli & Cabitza, 2024; Scannell et al., 2024). Some of the main recommendations are as follows: having independent ethics boards to review new technologies before they are implemented, strict rules for using AI in critical areas like healthcare, and independent reviews of data protection measures and representational fairness prior to system deployment in the market. Thus, there is a need for strong regulatory oversight in order to make these technologies to be ethical, safe, and fair.

As emotion-recognition technologies are being applied in consumer applications, regulatory frameworks must guarantee better user rights and protections (Scannell et al., 2024). Without these safeguards, the prospect of having emotional data exploited for commercial and political purposes remains high (Knight, 2024).

Table 1 summarizes the core ethical issues addressed in this paper and links them to the best practice of informed consent, privacy, fairness, bias, and user trust. This short summary gives clear and easily implementable guidelines for ethical practice in the conduct of emotion-recognition research to researchers and practitioners. In integrating these considerations into one easy to use table, the critical regions that need to be addressed consistently in order to achieve ethical integrity and public confidence in emotion-recognition technologies are made apparent.

Table 1. The figure presents a summary of basic ethical practices in emotion-recognition research, which includes proactive engagement, protective system design, and trust-centered human interaction.

Conceptual Framework Areas	Accuracy & Fairness	Bias Mitigation	User Trust & Participation
Informed consent			Banzon et al. (2024); Cartwright et al. (2024)
Privacy and data protection			Cartwright et al. (2024); Muir et al. (2019)
Bias and fairness	Banzon et al. (2024); Hazmoune and Bougamouza (2024)	Banzon et al. (2024); Hazmoune and Bougamouza (2024)	
Psychological impact on participants			Muir et al. (2019); Schrank et al. (2014)
Dual-use concerns and misuse	Banzon et al. (2024); Cartwright et al. (2024)		Banzon et al. (2024); Cartwright et al. (2024)
Transparency and accountability	Cartwright et al. (2024); Muir et al. (2019)		Cartwright et al. (2024); Muir et al. (2019)
Cultural and contextual sensitivity	Banzon et al. (2024); Hazmoune and Bougamouza (2024)	Banzon et al. (2024); Hazmoune and Bougamouza (2024)	

Table 1. *Cont.*

Conceptual Framework Areas	Accuracy & Fairness	Bias Mitigation	User Trust & Participation
Commercialization and profit-driven motives			Cartwright et al. (2024)
Participant control over the data			Cartwright et al. (2024); Muir et al. (2019)
Regulatory compliance and institutional oversight	Banzon et al. (2024); Muir et al. (2019)		Banzon et al. (2024); Muir et al. (2019)

4. Methodology

Our methodology is guided by three primary hypotheses. First, we believe that emotion-recognition systems that use multimodal information (facial expression, voice, and physiological signals) are more accurate and equitable than unimodal systems (Hazmoune & Bougamouza, 2024). Second, including culturally rich datasets decreases algorithmic bias and therefore improves equitable outcomes for various demographic groups (Hazmoune & Bougamouza, 2024). Third, we believe that publicly available and easily understandable consent mechanisms enhance participants' trust and compliance in emotion-recognition research (Cartwright et al., 2024). In line with these hypotheses, we developed our methodological approach to address the ethical challenges identified in the previous sections of the paper in a systematic manner.

4.1. Literature Selection and Review Strategy

The research used a narrative conceptual review method to analyze the data. A peer-reviewed literature search was performed across databases Scopus, IEEE Xplore, PubMed, and Google Scholar. The research used the following search terms: emotion recognition, affective computing, ethics in AI, privacy in emotion AI, algorithmic bias in emotion detection, and emotion AI and consent. The selection of papers depended on their ethical focus on emotion-recognition technologies and their coverage of multiple application fields including education and healthcare and surveillance and their conceptual insights into fairness and transparency and informed consent and data protection. The review includes a wide range of influential recent studies from 2020 to 2025 together with essential foundational texts by Cowie et al. (2001); Picard (1997). The main emphasis was on recent literature but older sources were added when necessary to support conceptual foundations and historical background.

4.2. Conceptual Framework to Methodological Principles Mapping

The conceptual framework identifies ten ethical areas in emotion-recognition research, each linked to one or more of three core methodological principles: accuracy and fairness, bias mitigation, and user trust and participation. These mappings are detailed in Table 1 and visually represented in Figure 3, with each linkage supported by current literature.

All ten ethical areas are directly related to user trust and participation. Ethical domains such as informed consent, psychological impact, and cultural sensitivity show particularly strong alignment with this principle. Others—such as dual-use awareness, transparency, and regulatory compliance—span multiple methodological principles, demonstrating the interconnectedness of ethical challenges in this field. This layered distribution illustrates the complexity and interdependence of ethical considerations in the development and deployment of emotion-recognition systems.

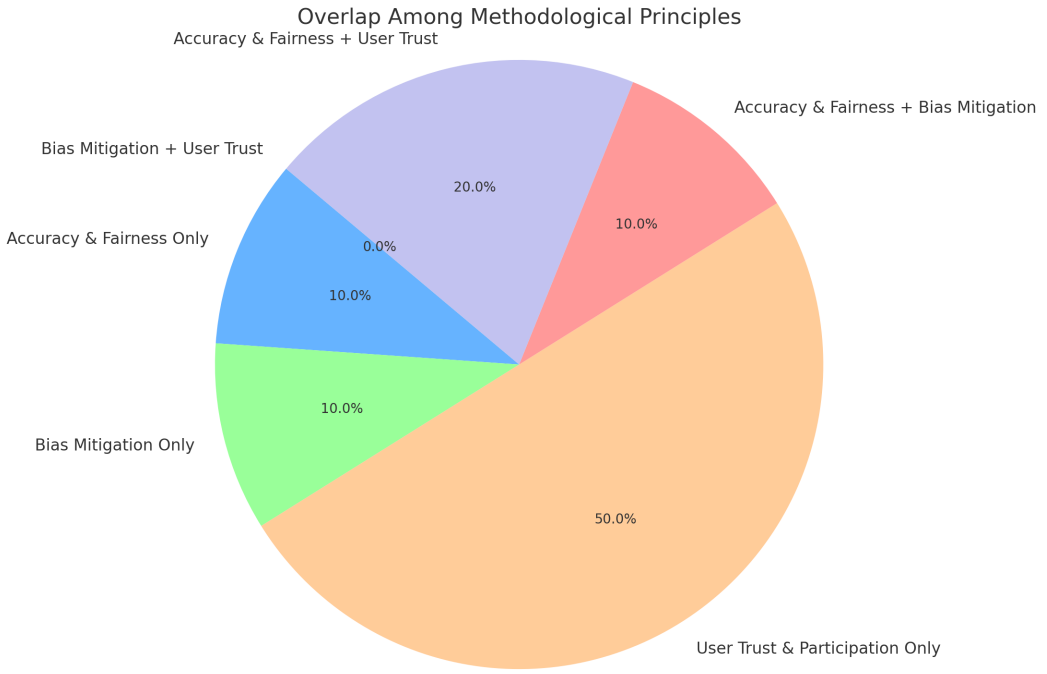


Figure 3. Methodological gaps in ethical emotion-recognition research.

Research indicates that emotion-recognition systems, which process multiple input modalities including facial expressions and voice and physiological signals achieve superior accuracy and fairness compared to systems that rely on single input modalities (Hazmoune & Bougamouza, 2024). It has been established that the use of culturally enriched datasets significantly minimizes bias in algorithms and leads to fair outcomes for different population groups (Banzon et al., 2024; Hazmoune & Bougamouza, 2024). According to Cartwright et al. (2024), open and easily understandable consent processes should enhance participants' confidence and cooperation in emotion-recognition studies.

This paper proposes a holistic approach to the ethics of designing, implementing, and deploying emotion-recognition systems, with a focus on fairness, accountability, and transparency. The proposed approach involves evaluating the ethical risks and possible solutions to the risks, which will be used in the deployment of emotion-recognition systems (Parraga et al., 2025; Scannell et al., 2024). At its core, our methodology suggests a practical checklist approach, based on ethical best practices, which can be used by researchers to guide the design and implementation of their studies (Lazar et al., 2017). The remaining subsections expand on the specific measures that can be taken to address these ethical issues effectively.

4.3. Transparency Through Interactive Dashboards

Transparent reporting and visualization contribute directly to accuracy, trust, and non-bias in performance monitoring processes (Cartwright et al., 2024; Muir et al., 2019). Such dashboards provide stakeholders with real-time insights into how these systems use and work with the data. The visualizations of the data flows and the algorithmic decision making are shown, such as how the inputs like facial expressions, vocal tones, or physiological signals are analyzed. This is especially important in educational settings because the transparency means that the tools for affect detection in classrooms are ethical and do not have provisions for violating students' privacy or autonomy (Banzon et al., 2024).

4.4. User Empowerment Through Consent and Data Management Tools

Empowerment mechanisms are consistent with ethical compliance by putting control in the hands of participants, thus contributing to bias mitigation through explicit, transparent data practices (Cartwright et al., 2024; Muir et al., 2019). It is therefore essential to empower users to make informed decisions about their participation for ethical compliance. Web-based consent forms provide simple and easy to read language of the information about the data collection and its use and other purposes for which the biometric data will be used (Behnke et al., 2024). The interactive applications allow the users to set their preferences and opt out of certain type of data usage or erasure of data. This is in consensus with the recommendations made by Behnke on informed consent and good data management in wearable technologies (Behnke et al., 2024). This way, through dynamic feedback and customization options, users are in control of their personal information and challenges to autonomy and misuse.

4.5. Stakeholder Education and Ethical Awareness

Increased ethical awareness helps to reduce biases and achieve more equitable algorithmic decisions by making stakeholders aware of the possible sources of biases and the effects of biases (Banzon et al., 2024; Cartwright et al., 2024; Hazmoune & Bougamouza, 2024). Educational tools are crucial in the promotion of ethical practices for developers, researchers, and end users. Cartwright argued that there is a need to include emotional intelligence and ethical thinking in the development of emotion-sensitive learning algorithms (Cartwright et al., 2024). These resources are developed with specific target groups in mind, so that the ethical principles are understood and easily applied.

4.6. Ethical Oversight Through Virtual Audits and Stakeholder Engagement

Ethical oversight mechanisms guarantee continuous conformity to fairness standards and the stakeholder engagement increases the ethical credibility of the research processes (Cartwright et al., 2024; Vigil et al., 2024). Mechanisms that check emotion-recognition systems' compliance with the relevant rules at all stages of the product life cycle are useful. Virtual ethical audits replicate real-world conditions to determine the possible risks and to assess the controls in place to mitigate such risks. As pointed out by Vigil, an interdisciplinary approach is critical for solving complex ethical issues that can occur in various application areas, including the use of emotion recognition for animal welfare (Vigil et al., 2024).

4.7. Continuous Monitoring and Compliance Mechanisms

Measures increase the confidence of the users and other stakeholders and also ensure that the rules of the game and the laws are well followed (Banzon et al., 2024; Muir et al., 2019). The mechanisms for ethical compliance should be built into the processes to enable the monitoring of ethical compliance in real time. The system performance and user feedback can be displayed on the ethical dashboards, and issues such as algorithmic biases or data breaches are addressed promptly. Using built-in features that send alerts to administrators to indicate possible ethical violations allow taking action immediately. According to Neethirajan, the inclusion of feedback loops in the system design improves trust and accountability as the system's ethical practices are updated in line with the technology (Neethirajan, 2024).

4.8. Implementation and Evaluation Strategy

Fairness, bias mitigation, user satisfaction, and regulatory compliance are the evaluation metrics for ethical implementation and achieve ethical consistency in various settings

to develop a clear, ethical framework (Hazmoune & Bougamouza, 2024; Parraga et al., 2025; Scannell et al., 2024). The proposed conceptual framework will be tested in practice during pilot studies in various application domains, including education, healthcare, and public safety. For instance, Hazmoune and Bougamouza highlighted the need of an unbiased and inclusive dataset for improving the performance of the multimodal emotion-recognition systems (Hazmoune & Bougamouza, 2024).

5. Potential Impact on Society

Using emotion-recognition systems that are both transparent and fair can help educators understand students' emotions which in turn can help in adapting learning strategies to the needs of diverse students (Banzon et al., 2024). However, the implementation of these systems is not without its challenges, and the ethical considerations include user understanding, equitable design, and accountability in emotion processing need to be addressed (Banzon et al., 2024; Cartwright et al., 2024).

Emotion-recognition systems are thus seen as potential tools for personalized diagnosis and treatment in healthcare settings, which are most useful in mental health care, for instance, in the management of emotional numbing and enhancement of therapeutic processes (Muir et al., 2019). However, such applications raise critical ethical issues including those to do with patient privacy, autonomy, and well-being (Muir et al., 2019; Rizzo et al., 2003).

Lastly, the more general social deployment, including safety, security, and surveillance uses, only stresses the importance of ethical oversight and accountability in the case of emotion-recognition technologies (Cartwright et al., 2024; Rizzo et al., 2003). Thus, continuous ethical oversight, transparent data management, and stakeholder engagement will be crucial for the ethical integration of emotion-recognition technologies into society.

6. Future Works

The recent transformer-based multimodal emotion-recognition approaches show promising results in combining different data types including facial, vocal, and physiological signals to improve both accuracy and demographic equity (Hazmoune & Bougamouza, 2024). Further research on these models may help to reduce bias across populations and support more realistic, inclusive deployment in real-world settings.

The current state of emotion-recognition research methodological convergence appears in Figure 3, which shows the relationships between predictive accuracy, demographic robustness, and stakeholder engagement. The 0% value in stakeholder engagement demonstrates an essential research practice deficiency because it shows that user trust and participatory design and real-world feedback mechanisms are currently absent from research methods. The absence of trust and user participation in practical methodological frameworks demonstrates an underlying equity deficit which affects both design and evaluation processes. Future research needs to focus on resolving these disparities especially those concerning inclusion and outcome consistency to create ethically resilient and socially aligned systems. The development of reliable emotion-recognition technologies requires standards that integrate these methodological dimensions.

Another important area is the improvement of culturally enriched datasets in order to increase the robustness of the algorithms. The reduction of the bias in the datasets with respect to the demographic factors including age, ethnicity, and cultural background will greatly enhance the accuracy and fairness of the emotion-recognition systems (Kissman & Maurer, 2002). In addition, this increased diversity will allow the researchers to design algorithms that are less likely to be biased and more likely to be effective in a variety of emotions and environments (Behnke et al., 2024).

The application of ethical frameworks in real conditions also needs further investigation. The pilot studies in educational and healthcare organizations can help to concretize the assessment of algorithmic fairness, user satisfaction, and the ethical impacts of deploying these technologies (Hazmoune & Bougamouza, 2024; Schrank et al., 2014). Formulation of policy guidelines for the use of emotion-recognition technologies, such as wearables, will be useful in protecting the privacy, consent, and equity of the users (Behnke et al., 2024).

Emotion-recognition research and implementation are complex and the understanding of the ethical issues in this area is complex, thus contribution from psychology, computer science, ethics, and policy making is needed (Pavarini et al., 2021; Rizzo et al., 2003). This approach guarantees that the progress being made is holistic and sustainable and that it addresses the needs of society as a whole. In general, the examined emotion-recognition technologies have many advantages, but the right design and use of these technologies require further partnership, research, and reflection on ethics, power, and equity (Banzon et al., 2024; Jagtap et al., 2023; Thieme et al., 2015). Future research could build on the ethical issues discussed here by reviewing the ethical frameworks that have been proposed for virtual reality, in order to develop more comprehensive guidance for the new types of interactive technologies (Raja & Al-Baghli, 2025).

The new system-level emotion-recognition approaches of Harmonia (Anthireddygari & Barker, 2024) and Thelxinoë (Barker & Levkowitz, 2024a, 2024b) demonstrate how ethical safeguards can be built into the fundamental architecture of these systems. These projects demonstrate the need to link technological progress with changing ethical requirements in practical use cases.

7. Conclusions

This work proposes a unified ethical framework for emotion-recognition research, involving ten key considerations across the stages of design, implementation, and deployment of emotion-recognition systems as shown in Figure 2. At the level of the individual, the framework emphasizes knowledge and participant autonomy with regard to informed consent, so that individuals know and can control the use of their emotional data. It also guarantees privacy and data security by ensuring that emotional data is protected according to the data protection laws. The consideration of the psychological effects of these technologies is also included to avoid negative emotional effects or stress that may be incurred by the users.

On a broader system level, the framework recommends equity by addressing algorithmic bias, which is backed by more representative datasets that include a wide cross-section of populations (Kargarandehkordi et al., 2024). It advocates for the consideration of cultural factors in the development and use of emotion-recognition systems, as a way of respecting the fact that what is considered as emotional, and how it is expressed and interpreted, is culturally determined. The implementation of periodic ethical audits and compliance reviews after deployment is suggested by ongoing research on algorithmic fairness and system transparency (Hazmoune & Bougamouza, 2024).

The framework also raises concerns on the dual-use dilemma, where useful technologies can be reverted for negative purposes like surveillance or control (Elyoseph et al., 2024). To ensure the credibility of the public, emotion data should be collected, interpreted, and used openly and it should be possible to trace and attribute system decisions to their source. Additionally, the ethics of commerce are also discussed to ensure that the business vision does not overshadow the ethical principles and that vulnerable populations are not taken advantage of in the name of innovation (Elyoseph et al., 2024). Ultimately, a strong set of regulations is identified as crucial to guarantee the practice of these principles (Denecke & Gabarron, 2024). Therefore, these tenets can be regarded as a strong package

of safeguards against foreseeable misuse and harm, which will help to maintain a proper balance between the respect for human dignity and the development of technology.

The process of turning ethical principles into actual protective measures needs more than theoretical work; it requires coordinated participation that goes beyond academic environments. The EU AI Act serves as a foundation for policymakers to expand existing regulations because it already identifies emotion recognition as a high-risk category that needs strict monitoring (Denecke & Gabarron, 2024). The successful transition from ethical theory to operational practice depends on institutions implementing complete governance tools, which include dynamic participant agreements and built-in data safeguards and fairness auditing mechanisms and public-facing accountability measures.

The ethical integration process depends heavily on developers and industry leaders to lead its advancement. The design philosophy should be based on proactive responsibility, which includes user-facing control mechanisms and cultural sensitivity evaluations and rigorous testing for demographic reliability before deployment. Developers need to maintain ongoing disclosure regarding the complete processing and storage and interpretation procedures for affective data. A trustworthy development lifecycle requires technologists to understand their accountability through ethics reviews and stakeholder feedback loops and formal impact assessments at every stage of system design.

Hence, the cooperation of regulatory bodies and technology developers is important: policymakers provide the adaptive governance and management, while developers implement and sustain these values in the day to day design decisions. In this way, emotion-recognition technologies can be developed in a socially acceptable and humane way with built-in mechanisms to address ethical concerns when they appear.

For the future, this work also shows several key directions for further research of the ethical basis of emotion recognition. A major contribution in this regard is the generation of new, more representative and diverse datasets that include a wider range of subjects and emotional displays to increase the diversity of the samples used to train emotion-recognition models (Kargarandehkordi et al., 2024). This is also true for the need of longitudinal research to explore the lasting positive and negative psychological impacts on people who use emotion-sensing systems on a daily basis.

These would help to reveal any long term trends in effects on well being, trust, and behavior that may not appear in short term experiments, and thus reveal thresholds for safe use and application. Researchers should focus on the real-life testing of the proposed ethical framework by applying it to the pilot projects of emotion-recognition systems. Applying these ethical principles in practice—in schools, workplaces, healthcare, or other real life settings—will help to reveal the specific problems and will help to adapt the guidelines based on the actual experience of how the harm can be avoided.

This type of evaluation is important for transferring theory into practice and moving beyond simply stating good practice. Finally, there is a critical need for adaptive regulation that is able to keep up with the fast pace of development of emotion AI. Regulatory and oversight bodies should be reviewed and revised from time to time in the light of new technologies and experiences in the field (Denecke & Gabarron, 2024).

The proposed framework needs to be adaptable to accommodate regular updates of new technologies while maintaining built-in safeguards and fairness-enhancing capabilities. Global coordination is essential, as many emotion-recognition systems are deployed across borders and require consistent oversight. To prevent regulatory blind spots, international bodies and national governments must align legal and ethical standards with emerging technological developments through continuous review and proactive policy adaptation.

Thus, the ethical framework suggested in this paper, together with the proposed actions and future studies, can be considered as a step by step guide to the ethical use

of emotion-recognition technologies. By anticipating these principles, policymakers and developers can reduce the ethical risks that currently exist and are likely to arise in the future, and thereby promote public confidence in emotion AI. The continued focus on ethical design and management of these innovations in emotion recognition will make sure that the developments in this field are contributing to the public benefit and upholding people's rights and dignity, thus ensuring that the technology in question is evolving in tune with the needs and expectations of the people it affects. The ethical framework should stay flexible while undergoing periodic assessments for updates that address new technological challenges and applications (Raja & Al-Baghli, 2025). The framework provides a path to turn ethical theory into measurable practice through its integration of system design and governance principles that include accuracy and bias mitigation and user trust.

Author Contributions: Conceptualization, D.B., B.A. and M.K.R.T.; methodology, D.B. and B.A.; software, D.B., and A.F.; validation, D.B. and A.F.; formal analysis, D.B. and B.A.; investigation, D.B., M.K.R.T., A.F. and B.A.; resources, B.A. and D.B.; data curation, D.B.; writing—original draft preparation, D.B. and M.K.R.T.; writing—review and editing, D.B., M.K.R.T. and A.F.; visualization, D.B.; supervision, B.A.; project administration, D.B. and B.A.; funding acquisition, B.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used and analyzed in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: Author Darlene Barker was employed by the company DB Attic Studios, LLC. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Anthireddygari, A. R., & Barker, D. (2024). Harmonia: Development of a standardized database for emotion recognition research. In *Lecture notes in electrical engineering*. Springer.

Banzon, A. M., Beever, J., & Taub, M. (2024). 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. [CrossRef]

Barker, D., & Levkowitz, H. (2024a). Thelxinoë: Recognizing emotions using gradient boosting, LSTM, and EEG. In *Lecture notes in electrical engineering*. Springer.

Barker, D., & Levkowitz, H. (2024b). Thelxinoë: Recognizing human emotions using pupillometry and machine learning. *Machine Learning and Applications: An International Journal*, *11*(1), 1. [CrossRef]

Behnke, M., Saganowski, S., Kunc, D., & Kazienko, P. (2024). Ethical considerations and checklist for affective research with wearables. *IEEE Transactions on Affective Computing*, *15*(1), 50–62. [CrossRef]

Benouis, M., Andre, E., & Can, Y. S. (2024). Balancing between privacy and utility for affect recognition using multitask learning in differential privacy-added federated learning settings: Quantitative study. *JMIR Mental Health*, *11*, e60003.

Burgio, F., Menardi, A., Benavides-Varela, S., Danesin, L., Giustiniani, A., Van den Stock, J., De Mitri, R., Biundo, R., Meneghello, F., Antonini, A., Vallesi, A., de Gelder, B., & Semenza, C. (2024). Facial emotion recognition in individuals with mild cognitive impairment: An exploratory study. *Cognitive, Affective, & Behavioral Neuroscience*, *24*, 599–614. [CrossRef]

Cartwright, J., Ellington, M. T., & Hawthorne, A. S. (2024). Exploring the potential of emotional intelligence in AI through emotion-sensitive learning algorithms. *Preprint*. Available online: https://www.researchgate.net/publication/378288861 (accessed on 24 March 2025).

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, *18*(1), 32–80. [CrossRef]

Denecke, K., & Gabarron, E. (2024). The ethical aspects of integrating sentiment and emotion analysis in chatbots for depression intervention. *Frontiers in Psychiatry*, *15*, 1462083. [CrossRef]

Di Dario, D., Pentangelo, V., Colella, M. I., Palomba, F., & Gravino, C. (2024). Collecting and implementing ethical guidelines for emotion recognition in an educational metaverse. In *Adjunct proceedings of the 32nd ACM conference on user modeling, adaptation and personalization* (pp. 549–554). Association for Computing Machinery.

Dominguez-Catena, I., Paternain, D., & Galar, M. (2024). Metrics for dataset demographic bias: A case study on facial expression recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(8), 5209–5226. [CrossRef]

Durachman, Y., Putra, S. J., Nanang, H., & Sukmana, H. T. (2024, August 7–8). *Analysis sentiment of public opinion on social media using naïve Bayes and TF-IDF algorithms*. 2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT) (pp. 1–6), Tangerang, Indonesia. [CrossRef]

Elsey, J., & Kindt, M. (2016). Manipulating human memory through reconsolidation: Ethical implications of a new therapeutic approach. *AJOB Neuroscience*, *7*(4), 225–236. [CrossRef]

Elyoseph, Z., Hadar Shoval, D., & Levkovich, I. (2024). Beyond personhood: Ethical paradigms in the generative artificial intelligence era. *The American Journal of Bioethics*, *24*(1), 57–59. [CrossRef]

Florestiyanto, M. Y., Surjono, H. D., & Jati, H. (2024). Harnessing emotions for effective online learning: A systematic review of emotion recognition techniques. In A. P. Suryotomo, & H. C. Rustamaji (Eds.), *Proceedings of the 2023 1st international conference on advanced informatics and intelligent information systems (ICAI3S 2023)* (pp. 72–87). Atlantis Press. [CrossRef]

Han, D., Park, H., & Rhee, S.-Y. (2021). The role of regulatory focus and emotion recognition bias in cross-cultural negotiation. *Sustainability*, *13*(5), 2659. [CrossRef]

Hazmoune, S., & Bougamouza, F. (2024). Using transformers for multimodal emotion recognition: Taxonomies and state of the art review. *Engineering Applications of Artificial Intelligence*, *133*(Part D), 108339. [CrossRef]

IEEE. (2024). *IEEE standard for ethical considerations in emulated empathy in autonomous and intelligent systems* (IEEE Std 7014-2024, pp. 1–51). IEEE. [CrossRef]

Jagtap, S., Marne, A., Sheikh, A., Potdar, V., & Chate, P. (2023). Design and analysis of unlocking student emotions: Enhancing e-learning with facial emotion detection. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, *11*(12), 1824–1830. [CrossRef]

Jindal, V., & Singh, K. (2025). Emotion recognition in AI: Bridging human expressions and machine learning. *International Journal For Multidisciplinary Research (IJFMR)*, *7*(1), 1–10. [CrossRef]

Jogi, S., Warang, S., Bhor, H. N., Solanki, D., & Patanwadia, H. (2024, April 19–20). *NLP unleashed: Transforming employment landscapes with dynamic recruitment platforms*. 2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT) (pp. 455–459), Sonepat, India. [CrossRef]

Kaiser, B. N., Varma, S., Carpenter-Song, E., Sareff, R., Rai, S., & Kohrt, B. A. (2020). Eliciting recovery narratives in global mental health: Benefits and potential harms in service user participation. *Psychiatric Rehabilitation Journal*, *43*(2), 111–120. [CrossRef]

Kargarandehkordi, A., Kaisti, M., & Washington, P. (2024). Personalization of affective models using classical machine learning: A feasibility study. *Applied Sciences*, *14*(4), 1337. [CrossRef]

Khan, U. A., Xu, Q., Liu, Y., Wang, J., Zhang, C., & Li, H. (2024). Exploring contactless techniques in multimodal emotion recognition: Insights into diverse applications, challenges, solutions, and prospects. *Multimedia Systems*, *30*, 115.

Kissman, K., & Maurer, L. (2002). East meets West: Therapeutic aspects of spirituality in health, mental health and addiction recovery. *International Social Work*, *45*(1), 35–43. [CrossRef]

Kleinrichert, D. (2024). Empathy: An ethical consideration of AI & others in the workplace. *AI & Society*, *39*, 2743–2757. [CrossRef]

Knight, W. (2024). This new tech puts AI in touch with its emotions—and yours. *WIRED*. Available online: https://www.wired.com/story/hume-ai-emotional-intelligence/ (accessed on 24 March 2025).

Kotian, A. L., Nandipi, R., M, U., S, U. R., VARSHAUK, & T, V. G. (2024, January 4–6). *A systematic review on human and computer interaction*. 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) (pp. 1214–1218), Bengaluru, India. [CrossRef]

Lakkaraju, K., Srivastava, B., & Valtorta, M. (2024). Rating sentiment analysis systems for bias through a causal lens. *IEEE Transactions on Technology and Society*, *5*(1), 82–92. [CrossRef]

Lal, M., & Neduncheliyan, S. (2024). Enhanced V-Net approach for the emotion recognition and sentiment analysis in the healthcare data. *Multimedia Tools and Applications*, *83*, 72765–72787. [CrossRef]

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction* (2nd ed.). Morgan Kaufmann.

Lin, W., & Li, C. (2023). Review of studies on emotion recognition and judgment based on physiological signals. *Applied Sciences*, *13*(4), 2573. [CrossRef]

Mattioli, M., & Cabitza, F. (2024). Not in my face: Challenges and ethical considerations in automatic face emotion recognition technology. *Machine Learning and Knowledge Extraction*, *6*(4), 2201–2231.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1–35.

Morellini, N. A. (2022). *The emotional brain: Fundamentals, practice, and innovation*. Springer.

Muir, H. J., Coyne, A. E., Morrison, N. R., Boswell, J. F., &Constantino, M. J. (2019). Ethical implications of routine outcomes monitoring for patients, psychotherapists, and mental health care systems. *Psychotherapy*, *56*(4), 459–469. [CrossRef]

Narimisaei, J., Naeim, M., Imannezhad, S., Samian, P., & Sobhani, M. (2024). Exploring Emotional Intelligence in Artificial Intelligence Systems: A Comprehensive Analysis of Emotion Recognition and Response Mechanisms. *Annals of Medicine and Surgery*, (86), 84657–4663.

Neethirajan, S. (2024). From predictive analytics to emotional recognition: The evolving landscape of cognitive computing in animal welfare. *Computational and Structural Biotechnology Journal*, *23*, 100–110. [CrossRef]

Ng, C. H. (1997). The stigma of mental illness in Asian cultures. *Australian and New Zealand Journal of Psychiatry*, *31*(3), 382–390. [CrossRef]

Obaigbena, A., Lottu, O. A., Ugwuanyi, E. D., Jacks, B. S., Sodiya, E. O., Daraojimba, O. D., & Lottu, O. A. (2024). AI and human-robot interaction: A review of recent advances and challenges. *GSC Advanced Research and Reviews*, *18*(2), 321–330. [CrossRef]

Parraga, O., More, M. D., Oliveira, C. M., Gavenski, N. S., Kupssinskü, L. S., Medronha, A., Moura, L. V., Simões, G. S., & Barros, R. C. (2025). Fairness in deep learning: A survey on vision and language research. *ACM Computing Surveys*, *57*(6), 1–40. [CrossRef]

Pavarini, G., Smith, L. M., Shaughnessy, N., Mankee-Williams, A., Thirumalai, J. K., Russell, N., & Bhui, K. (2021). Ethical issues in participatory arts methods for young people with adverse childhood experiences. *Health Expectations*, *24*(5), 1557–1569. [CrossRef]

Picard, R. W. (1997). Affective computing. MIT Press.

Raja, U. S., & Al-Baghli, R. (2025). Ethical concerns in contemporary virtual reality and frameworks for pursuing responsible use. *Frontiers in Virtual Reality*, *6*, 1451273. [CrossRef]

Rizzo, A., Schultheis, M. T., & Rothbaum, B. O. (2003). Ethical issues for the use of virtual reality in the psychological sciences. In S. S. Bush & M. L. Drexler (Eds.), *Ethical issues in clinical neuropsychology* (pp. 243–280). Swets & Zeitlinger Publishers.

Scannell, B., Moore, L., & Hayes, R. (2024). *The time to (ai) act is now: A practical guide to emotion recognition systems under the ai act.* Available online: https://www.williamfry.com/knowledge/the-time-to-ai-act-is-now-a-practical-guide-to-emotion-recognition-systems-under-the-ai-act/ (accessed on 3 March 2025).

Schrank, B., Brownell, T., Tylee, A., & Slade, M. (2014). Positive psychology: An approach to supporting recovery in mental illness. *East Asian Archives of Psychiatry*, *24*(3), 95–103. [CrossRef]

Sidana, S., Thakur, S. N., Kumar, S., Gomathi, R. D., Sinha, A., Saini, A. S., Raj, P., Dixit, K., Kumar, V., & Al-khayyat, A. (2024, June 24–28). *Mitigating racial bias in facial expression recognition using enhanced data diversification techniques*. 15th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1–6), Kamand, India. [CrossRef]

Singh, V., & Kamboj, S. (2024). Exploring changes in public opinion, political campaigns, and political behavior through data science and machine learning. *IEEE Access*, *12*, 49686–49700. [CrossRef]

Smith, C. H., Molka-Danielsen, J., Webb-Benjamin, J.-B., & Rasool, J. (2025). The challenges of consent in a decentralised metaverse: Exploring ethically informed protections and standards to safeguard humans. *Frontiers in Virtual Reality*, *6*, 1401073. [CrossRef]

Sukumaran, A., & Manoharan, A. (2024). Multimodal engagement recognition from image traits using deep learning techniques. *IEEE Access*, *12*, 25228–25244. [CrossRef]

Thieme, A., Wallace, J., Meyer, T. D., & Olivier, P. (2015, July 13–17). *Designing for mental wellbeing: Towards a more holistic approach in the treatment and prevention of mental illness*. British HCI 2015—Proceedings of the 2015 British HCI Conference (pp. 1–10), Lincoln, UK. [CrossRef]

Vigil, M. S. A., Anand, A. K., Singh, J. P., & Misra, S. K. (2024). Emotion recognition of animals using natural language processing. *Journal of Advanced Zoology*, *45*(1), 1233–1245. [CrossRef]

Vlek, R. J., Steines, D., Szibbo, D., Kübler, A., Schneider, M.-J., Haselager, P., & Nijboer, F. (2012). Ethical issues in brain-computer interface research, development, and dissemination. *Journal of Neurologic Physical Therapy*, *36*(2), 94–99. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.