Ethical Considerations in Artificial Intelligence: A Comprehensive Disccusion from the Perspective of Computer Vision

Ni Li *

Bank of Suzhou, 215144 Suzhou, China

Abstract. This paper delves deeply into the multifaceted ethical challenges within the realm of computer vision, focusing intently on various ethical dimensions inherent in this cutting-edge field. It emphasizes the pressing need to address ethical concerns related to AI technologies, including algorithmic fairness, informed consent, public engagement, robust privacy protocols, transparency, and the integration of human judgment through human-in-the-loop systems. The study underscores the vital importance of collaboration among diverse stakeholders, including governments, businesses, academia, and society, to promote responsible and equitable AI practices within computer vision.Through meticulous examination, the paper highlights the urgency of balancing technological advancement with ethical considerations. It advocates for the development and implementation of ethical principles, ensuring that AI technologies align with societal values and promote fairness, transparency, and accountability. The collaborative efforts among various sectors are crucial to fostering an ethical framework that guides the responsible deployment of AI in the field of computer vision. By integrating ethical consciousness into the core of technological innovation, this approach aims to create a symbiotic relationship between artificial intelligence and society, ultimately benefiting humanity as a whole.

1.Introduction

The rapid advancement of Artificial Intelligence (AI) technologies, particularly in computer vision, has not only transformed various sectors but also sparked significant ethical concerns[1]. As AI systems integrate further into our lives, questions about their ethical implications have come to the forefront. Issues like data privacy, algorithmic bias, and concentrated power have become central topics in AI ethics discussions, covering concerns from data misuse to potential threats to democratic principles[2, 3].

In response, stakeholders, including the scientific community, industry leaders, and policymakers, collaborate to navigate these complex ethical challenges. Strategies and frameworks are developed to ensure ethical alignment in AI development, deployment, and application. The media and the public, aware of AI's transformative potential, demand transparency and inclusivity in these efforts[4].

This paper conducts an in-depth exploration of AI ethics, focusing specifically on computer vision's ethical dimensions. Emphasizing ethical considerations over legal aspects, the study delves into the intricate moral challenges presented by AI technologies. It stresses the urgency of addressing these concerns and underscores the importance of collaboration among diverse stakeholders to promote responsible and just AI practices within computer vision. Subsequent sections critically examine these ethical challenges, emphasizing the interconnected nature of ethics and AI's future within the domain of computer vision.
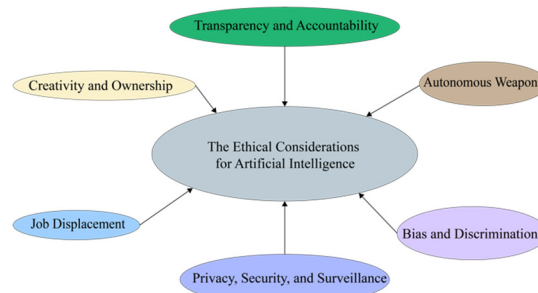
2.Ethics in Artificial Intelligence

Fig. 1. Challenges in AI Ethics.

2.1.Ethical Considerations in AI

Fig. 1 illustrates a multitude of challenges confronting AI ethics. Firstly, AI systems, embedded within vast datasets, can perpetuate societal biases, leading to unfair outcomes in pivotal areas like hiring, lending, and criminal justice. Secondly, AI systems often operate opaquely, particularly in critical sectors such as healthcare and autonomous vehicles, posing transparency and interpretability challenges.

Additionally, the creation of digital art through AI raises ownership concerns, especially when generated using others' AI systems. Persistent issues like fake news and social manipulation persist, with AI algorithms

---

* Corresponding author: lini2024@163.com

potentially exacerbating social divisions and interfering with elections. Expanding AI raises concerns about personal privacy, security, and surveillance, necessitating robust safeguards.

Furthermore, AI automation threatens widespread unemployment, while AI-powered autonomous weapons raise ethical alarms, prompting the need for international agreements and regulations.

2.2.Ethical Imperatives in AI

In an era where Artificial Intelligence (AI) pervades our daily lives, the ethical oversight of AI technology becomes paramount. AI systems, capable of amplifying biases and privacy concerns, introduce complex ethical challenges. Issues surrounding transparency, debates on job displacement, and global disparities in AI development exacerbate these ethical dilemmas. Effectively navigating these challenges requires interdisciplinary collaboration among technologists, ethicists, and policymakers. Addressing these ethical complexities necessitates cooperation between technical experts, policymakers, ethicists, and society at large. Establishing comprehensive regulations, transparent AI frameworks, diverse development practices, and fostering ongoing dialogues are crucial for ensuring responsible AI deployment becomes the norm. Initial efforts in AI governance concentrated on ethical frameworks, but practical implementation faced hurdles due to the absence of universal norms and clear methodologies. Integrating ethical considerations, particularly regarding biases and transparency, holds pivotal importance in both AI development and deployment.

Ethical governance mandates proactive measures, including transparent decision-making processes and active engagement with diverse stakeholders. Ethical guidelines must go beyond symbolic gestures, ensuring concrete and accountable practices within the AI industry. Striking a delicate balance between innovation and ethical principles, especially concerning AI's societal impacts, requires continuous ethical scrutiny and deliberate decision-making. Collaborative initiatives are indispensable to align AI with ethical standards, fostering fairness, inclusivity, and societal well-being.

3.Ethical Principles in Computer Vision

In the ever-progressing realm of computer vision, ethical considerations serve as the cornerstone, guiding the development and deployment of this transformative technology. Ethical approaches in computer vision involve proactive measures, responsible practices, and ongoing scrutiny. The six ethical principles of computer vision are shown in Fig. 2.

Algorithmic fairness, the first tenet, emphasizes the necessity of diverse and unbiased training datasets, requiring constant collaboration between ethicists and data scientists to rectify biases continuously. Informed consent, the second pillar, underscores the importance of upholding individual autonomy, particularly in scenarios involving facial recognition or image analysis, demanding

clear explanations about data usage to empower individuals in decision-making processes.

Public engagement, the third facet, advocates for active involvement of communities impacted by computer vision technologies. Through mechanisms like town hall meetings and public forums, the technology aligns with societal values, fostering transparency and inclusivity. Robust privacy protocols, the fourth principle, safeguard individuals' privacy rights through encryption and stringent access controls, ensuring data usage adheres strictly to specified purposes and durations, establishing trust in the technology.

Explainability and transparency, the fifth element, address the opacity of computer vision algorithms. Encouraging the development of interpretable models enhances understanding among end-users, fortifying trust between technology and society. Human-in-the-loop systems, the final dimension, introduce human oversight, offering nuanced judgment and ethical integrity, thereby refining the decision-making processes.

This comprehensive ethical framework, weaving together fairness, consent, public engagement, privacy, transparency, and human oversight, embodies the essence of responsible computer vision practices. This dynamic approach ensures that ethical considerations evolve alongside societal values and technological progress, nurturing a symbiotic relationship between technology and society, where trust, transparency, and fairness prevail, enriching the ethical discourse in this ever-advancing field. In the chapters that ensue, each of these ethical principles will be dissected meticulously. Through rigorous analysis and real-world illustrations, we illuminate the complexities surrounding these pillars, paving the way for ethically grounded advancements in the dynamic realm of computer vision.
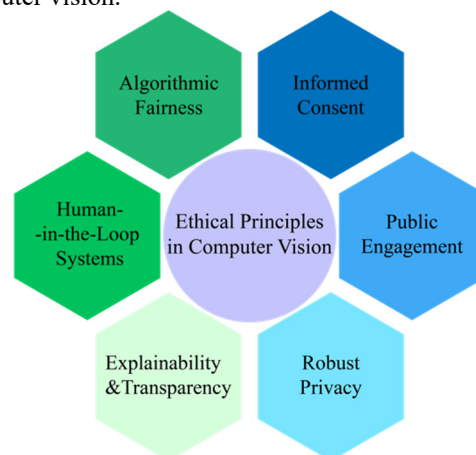

Fig. 2. Ethical Principles in Computer Vision.

3.1.Algorithmic Fairness

Ensuring fairness in computer vision algorithms is of utmost importance[5]. AI systems must be trained on diverse and representative datasets to prevent biases. Collaboration between ethicists and data scientists is crucial to continuously identify and rectify biases. The pursuit of algorithmic fairness is an ongoing commitment, evolving alongside societal understanding.

Firstly, diverse and representative datasets are essential for training fair computer vision algorithms. If the dataset leans towards specific demographics, the algorithm might exhibit biases, leading to unfair outcomes. Therefore, collecting widely varied data representing different races, genders, ages, cultures, and social backgrounds is key to ensuring algorithmic fairness.

Secondly, collaboration between ethicists and data scientists is indispensable. Ethicists can provide ethical guidance, aiding in identifying potential biases and ensuring algorithms do not unfairly impact specific demographics. Data scientists, on the other hand, are responsible for analyzing data, ensuring the algorithm's performance is fair across various demographics. Their collaboration helps uncover and rectify biases within algorithms, ensuring transparency and fairness for all populations.

Most importantly, algorithmic fairness is a continually evolving field. As societal understanding of fairness deepens, algorithms need constant adjustment and improvement. Therefore, continuous commitment and effort are vital to ensuring algorithmic fairness. Only through continuous learning, improvement, and adaptation can the field of computer vision achieve genuine fairness and equity.

3.2. Informed Consent

Upholding individual autonomy is a core principle in ethical computer vision practices. It means respecting people's right to control their own information and make decisions about how their data is used[6]. In scenarios where facial recognition or image analysis is involved, obtaining informed consent from individuals is not just important but essential. Ethical guidelines must require clear and straightforward explanations, outlining precisely how individuals' data will be utilized. These guidelines should be explicit and easily understandable, ensuring that individuals are fully aware of the purposes for which their data is being collected and processed.

Empowering individuals with comprehensive knowledge is crucial in this process. By providing them with detailed information about how their data will be used, the potential risks involved, and their rights regarding their personal information, they are better equipped to make informed decisions. This empowerment enables active participation, allowing individuals to voice concerns, set boundaries, and contribute to shaping the ethical landscape of computer vision technology.

In essence, ethical computer vision practices not only involve obtaining consent but doing so in a transparent, clear, and comprehensible manner. This approach respects individuals' autonomy, promotes informed decision-making, and fosters a relationship of trust between technology developers and the people whose data is being processed.

3.3. Public Engagement

Ethical practices in computer vision require more than just technical expertise; they demand active engagement with the broader public. Communities directly affected by the implementation of these technologies must be given the opportunity to express their concerns, expectations, and opinions[7, 8]. Their perspectives are invaluable in shaping the ethical framework that governs the development and deployment of computer vision systems.

To facilitate this engagement, various mechanisms such as town hall meetings, public forums, and online feedback platforms play a crucial role. Town hall meetings provide a physical space where community members can interact with experts, ask questions, and express their concerns face-to-face. Public forums, both online and offline, serve as platforms for open discussions, allowing diverse voices to be heard and different viewpoints to be considered. Additionally, feedback platforms enable continuous communication, allowing the public to provide input on ongoing projects and voice their opinions on the ethical implications of specific applications.

By integrating these mechanisms into the decision-making processes, computer vision developers can ensure that their technology aligns seamlessly with societal values and expectations. Public participation not only promotes transparency by involving the community in the decision-making processes but also enriches the ethical discourse. Diverse perspectives and experiences bring depth to discussions, fostering a more inclusive understanding of ethical considerations in computer vision.

In summary, active engagement with the wider public is not just a best practice; it is a fundamental necessity in the ethical development of computer vision technology. By embracing public input, developers can create systems that are not only technically proficient but also ethically robust, reflecting the values and concerns of the communities they serve.

3.4. Robust Privacy Protocols

Protecting the privacy of individuals captured in visual data is a fundamental ethical imperative in the realm of computer vision[9]. This principle entails a set of specific practices and protocols aimed at preserving individuals' personal information and ensuring their data is used responsibly and ethically.

Encryption: One of the key measures in safeguarding privacy is encryption. Visual data should be encrypted during storage and transmission, ensuring that even if the data is intercepted, it remains inaccessible and secure. Encryption serves as a robust barrier against unauthorized access, enhancing the overall privacy protection.

Anonymization: Another crucial practice is anonymizing visual data. Anonymization techniques modify the data in a way that individuals cannot be readily identified. This process removes or encrypts personally identifiable information, such as faces or other unique features, making it extremely challenging to trace the data back to specific individuals. Anonymization helps in balancing the utility of data for research or analysis while preserving individuals' privacy.

Stringent Access Controls: Implementing stringent access controls is essential. Only authorized personnel with a legitimate need should have access to visual data.

Access permissions should be carefully regulated, ensuring that individuals who handle the data are trained in ethical practices and understand the importance of privacy. Regular audits and monitoring can help enforce these access controls effectively.

Ethical Guidelines and Standards: Ethical guidelines play a pivotal role in ensuring privacy protection. These guidelines should enforce rigorous standards regarding the collection, storage, and usage of visual data. Clear policies must be established, specifying the purposes for which the data is collected and the duration for which it will be retained. Ethical standards should be communicated transparently, both to the individuals whose data is being collected and to the personnel handling the data.

Limited Data Usage: One of the foundational principles is limiting data usage strictly to specified purposes and durations. Data should only be used for the purpose it was collected for and should not be repurposed without explicit consent. Additionally, data should be retained only for as long as necessary, and once the purpose is fulfilled, it should be securely deleted or anonymized.

By upholding these protocols and adhering to ethical guidelines, computer vision developers demonstrate a commitment to respecting individuals' privacy rights. This commitment fosters a sense of trust among the public, assuring them that their personal information is handled with care and integrity, thereby encouraging responsible adoption of this transformative technology.

3.5.Explainability and Transparency

Computer vision algorithms are frequently seen as mysterious "black boxes," meaning that their internal workings and decision-making processes are obscure and difficult to comprehend[10]. This opacity raises ethical concerns, especially when these algorithms impact people's lives and decisions. Ethical approaches in artificial intelligence (AI) emphasize the need for transparency, advocating for the development of AI systems that are interpretable and understandable.

Researchers and developers are not only encouraged but are morally obliged to create models that can be interpreted and explained. Interpretable models provide clear insights into how AI systems arrive at their conclusions, shedding light on the intricate processes hidden within these algorithms. By making these systems transparent, end-users and stakeholders gain comprehensible insights into the factors influencing AI-derived decisions, making the technology more accountable and trustworthy.

Transparency and explainability are not just ethical imperatives; they are essential components that strengthen the bond of trust between technology and society. When people can understand how AI systems work and how decisions are made, they are more likely to trust and accept the technology. Transparent AI systems also allow for effective oversight, enabling experts, regulators, and the general public to assess the fairness and reliability of these algorithms.

In practical terms, creating interpretable models involves developing algorithms that produce not only accurate but also understandable results. This may involve using techniques that provide explanations for the decisions made by AI systems, such as generating human-readable descriptions or visualizations of the processes involved. Additionally, documenting the development process and ensuring that decision-making criteria are clear and unbiased contribute to the transparency of these systems.

By prioritizing transparency and explainability, ethical AI practices pave the way for responsible deployment of computer vision technology. When society can comprehend and scrutinize the functioning of AI algorithms, it promotes accountability, fairness, and ultimately, trust in these powerful technologies.

3.6.Human-in-the-Loop Systems

Integrating human oversight into computer vision applications stands as a critical ethical practice. Human reviewers possess unique abilities to assess nuanced and ambiguous situations, skills that AI often lacks. This collaboration between human judgment and artificial intelligence elevates both the accuracy and ethical integrity of computer vision systems.

Human reviewers excel in interpreting complex scenarios, especially those involving subtle cues or cultural nuances. Their contextual understanding allows them to make ethical decisions in situations where AI algorithms might misinterpret, ensuring accurate and responsible outcomes.

This human-AI partnership not only validates and refines AI-generated results but also places ethical considerations at the forefront. Human reviewers, equipped with empathetic discernment, weigh societal implications, privacy concerns, and ethical boundaries. They actively engage with the technology, questioning and challenging to align it with ethical standards and societal values.

Moreover, human oversight fosters continuous improvement. Human reviewers identify errors, biases, and ethical concerns, paving the way for iterative enhancements in algorithms. This iterative process guarantees that technology evolves in harmony with ethical norms, reinforcing trust between technology and society.

In summary, the integration of human judgment adds a vital layer of ethical depth to computer vision applications. It ensures accuracy, navigates complexities, and upholds ethical standards, fostering a responsible and trustworthy relationship between technology and society.

4.Ethical Practical Measures for Computer Vision Implementation

In the field of computer vision and image processing, practical measures for ethical AI implementation are crucial to address the complex ethical and legal challenges posed by these technologies. Before deploying any computer vision technology, conducting thorough ethical

impact assessments is imperative. These assessments should be integrated into the development process, evaluating potential risks related to privacy, discrimination, and data ethics. By identifying these risks early on, developers can proactively design algorithms and systems that mitigate these concerns, ensuring ethical standards are maintained.

4.1. Algorithmic Fairness and Informed Data Usage

In the field of computer vision, challenges related to algorithmic fairness and ethical use of data have emerged prominently. Biases concerning attributes like gender and race can lead to unfair outcomes, prompting researchers to focus on ensuring fairness across various attributes in computer vision systems.

To tackle these challenges, researchers have employed diverse strategies. They have developed techniques for fairness in image classification, including methods such as balanced data training, adversarial training, and re-weighting of training samples. Additionally, efforts have been directed towards mitigating biases associated with gender and race. Techniques like debiasing neural networks and fairness-aware information alignment have been introduced. Addressing bias in dataset construction, tools have been devised to detect and eliminate biases in visual datasets. Specific domains like facial recognition have seen the implementation of methods such as group adaptive classifiers and balanced feature learning to ensure fairness.

Implementation of these methodologies not only enhances the accuracy of computer vision systems but also guarantees fairness and justice among diverse groups. Looking ahead, ongoing research in this area is expected to address emerging challenges, ensuring the ethical and equitable application of computer vision technology.

4.2. Privacy and Transparency Integration

The integration of privacy and transparency is critical in various domains, particularly in the context of autonomous driving and artificial intelligence applications. Instances of evasion attacks in autonomous driving can lead to vehicles violating traffic rules or even causing accidents, emphasizing the importance of robust privacy measures. Similarly, in the medical field, the introduction of malicious data can significantly skew AI models, impacting medication suggestions for a substantial portion of patients. These vulnerabilities highlight the dire need for privacy and transparency integration in AI systems.

In cancer screening, AI models might achieve high accuracy, yet medical professionals might only agree with about half of the results due to the lack of reasoning and logic provided by the AI, emphasizing the necessity of transparent decision-making processes. Furthermore, deep neural networks lack robustness and can be susceptible to evasion attacks, leading to erroneous AI judgments and compromising business security. The complexity of these networks also leads to a lack of transparency and interpretability, potentially violating legal regulations like GDPR, and even causing issues related to fairness, accuracy, and accountability, underscoring the importance of transparent and explainable AI algorithms.

The vast and diverse datasets used in AI applications, such as those in autonomous driving, might lack comprehensive data security, leading to risks like data leaks, tampering, and misuse. Moreover, without proper control over the purpose of AI applications, there is a potential for misuse, emphasizing the need for robust privacy regulations and ethical guidelines. Biased or unfair judgments due to data quality problems also accentuate the significance of integrating privacy measures, ensuring fairness and accuracy in decision-making.

Insufficient knowledge and skills among developers and deployers might lead to misusing AI systems, risking security and privacy incidents. Additionally, AI synthesis technologies, while aiding those with speech difficulties, can be exploited by fraudsters for activities like voice-based scams, necessitating stringent privacy measures. Misjudgments in credit systems or facial recognition mismatches due to improper parameters emphasize the necessity for clear legal frameworks and regulations, defining the rights and responsibilities of all stakeholders, especially in domains like autonomous driving and law enforcement.

Lastly, instances where AI algorithms can unintentionally reveal sensitive information, such as patient genetic data, underpin the critical importance of integrating robust privacy protections. The discussion around legal regulations and responsibilities intensifies, especially in scenarios where autonomous vehicles are involved in accidents, raising debates on regulatory frameworks and legal liabilities. In conclusion, the integration of privacy and transparency measures is pivotal in mitigating these risks and ensuring the responsible development and deployment of AI technologies across various domains.

4.3. Engagement and Consent

The analysis of 16 national AI strategies within the scope of computer vision indicates that public engagement values and mechanisms are rarely emphasized. While strategies acknowledge public involvement, these mentions are often abstract and overshadowed by other concerns. This could reflect either superficial rhetoric or a belief that public participation conflicts with AI's efficiency and economic benefits. Limited evidence suggests a lack of "upstream public deliberation" to balance AI's ethical and social risks, especially at the national level. Mentioning abstract values may encourage participation in specific contexts, particularly in countries like Finland and Germany where such values are prominent. However, the consistent link between public engagement, professional ideals, and the private sector implies a normalization of AI, echoing criticisms of neoliberal approaches. National strategies often prioritize market-led solutions and individual autonomy, sidelining genuine public engagement. This pattern deserves focused attention within the realm of computer vision.

5.Ethics Countermeasures in CV

The rapid evolution of computer vision technology has ushered in a new era of possibilities, transforming industries and revolutionizing how we interact with the digital world. However, amid the excitement, ethical and legal concerns loom large, necessitating a thoughtful and comprehensive approach to ensure the responsible development and deployment of computer vision systems.

5.1.Privacy Protection and Data Security

In an age where personal images and videos are integral components of computer vision datasets, ensuring the utmost privacy protection is paramount. Advanced encryption algorithms, coupled with robust storage protocols, are imperative to secure sensitive visual data. Moreover, transparent data usage policies, communicated clearly to individuals, empower them with control over their data. Informed consent mechanisms, utilizing user-friendly interfaces and clear language, enhance awareness, enabling users to make conscious decisions about data sharing.

5.2.Fairness and Anti-Discrimination Algorithms

Facial recognition and similar technologies, while promising enhanced security, must not perpetuate societal biases. Anti-discrimination algorithms, constantly monitored and refined, are pivotal in ensuring fairness and accuracy. Collaborative efforts with social scientists and ethicists help in identifying nuanced biases, leading to algorithmic improvements. Additionally, continuous public engagement via feedback mechanisms aids in detecting discriminatory patterns, enabling timely corrections.

5.3.Data Ethics and Transparent Usage

The ethical use of personal data is at the core of responsible AI. Ethical guidelines, rooted in legal frameworks, govern the collection and usage of data for model training. Transparent data usage and sharing policies, prominently displayed, cultivate a culture of responsible data practices. Ethical audits, conducted periodically, assess adherence to these guidelines, ensuring the legitimacy and fairness of data utilization.

5.4.Transparency and Interpretability

The opacity of algorithms often poses challenges in public understanding. Encouraging the development of interpretable AI technologies becomes pivotal. Detailed explanations, in non-technical language, regarding algorithmic decisions are essential. Interactive public workshops, led by AI experts and ethics educators, decode complex algorithms for the general populace. Transparent AI fosters trust and bridges the knowledge gap, enabling a more informed society.

5.5.Ethical Education and Interdisciplinary Collaboration

Ethical education, an integral component of computer science curricula, goes beyond technical skills. It instills ethical reasoning, empathy, and societal awareness in future tech professionals. Interdisciplinary collaboration expands perspectives, involving not only ethicists but also legal scholars, psychologists, and policymakers. Collaborative research projects ensure a holistic approach, enriching the ethical discourse and paving the way for innovative, responsible technologies.

In the dynamic landscape of computer vision, ethical strategies are not static principles but evolving guidelines. Through continuous dialogue, research, and interdisciplinary collaborations, the fusion of technological innovation with ethical consciousness becomes possible. By embracing these multifaceted strategies, the field of computer vision can navigate the complexities of ethics, fostering a harmonious coexistence between artificial intelligence and society.

6.Conclusion

By thoroughly examining the ethical challenges in the realm of computer vision, this paper proposes a series of proactive measures. In safeguarding privacy, ensuring fairness, combating discrimination, addressing data ethics, and promoting transparency, the balance between technological advancement and social ethics is emphasized. The importance of transparency and interpretability is underscored, fostering trust and understanding between technology and society. Moreover, the paper highlights the value of education, cultivating tech professionals with ethical reasoning, and advocates for interdisciplinary collaboration, ensuring diverse perspectives are incorporated into ethical discussions. Through ongoing dialogue, research, and collaborative efforts, the fusion of technological innovation with ethical consciousness becomes attainable. Embracing these multifaceted strategies, the field of computer vision can navigate the complexities of ethics, fostering a harmonious coexistence between artificial intelligence and society.

References

1. F. Li, N. Ruijs, and Y. Lu, "Ethics & AI: A systematic review on ethical concerns and related strategies for designing with AI in healthcare," AI, 4(1), 28-53 (2022).

2. A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," Nature machine intelligence, 1(9), 389-399 (2019).

3. B. C. Stahl, J. Antoniou, M. Ryan *et al.*, "Organisational responses to the ethical issues of artificial intelligence," AI & SOCIETY, 37(1), 23-37 (2022).

4. M. K. Kamila, and S. S. Jasrotia, "Ethical issues in the development of artificial intelligence: recognizing the

risks," International Journal of Ethics and Systems, (2023).

5. L. Gustafson, C. Rolland, N. Ravi *et al.*, "FACET: Fairness in Computer Vision Evaluation Benchmark." 20370-20382.

6. B. Pickering, "Trust, but verify: informed consent, AI technologies, and public health emergencies," Future Internet, 13(5), 132 (2021).

7. V. C. Müller, "Ethics of artificial intelligence and robotics," (2020).

8. C. Wilson, "Public engagement and AI: A values analysis of national strategies," Government Information Quarterly, 39(1), 101652 (2022).

9. Y. Tian, S. Wang, J. Xiong *et al.*, "Robust and privacy-preserving decentralized deep federated learning training: Focusing on digital healthcare applications," IEEE/ACM Transactions on Computational Biology and Bioinformatics, (2023).

10. A. Fernandez-Quilez, "Deep learning in radiology: ethics of data and on the value of algorithm transparency, interpretability and explainability," AI and Ethics, 3(1), 257-265 (2023).