(REVIEW ARTICLE)

Ethical AI in practice: Balancing technological advancements with human values

Benjamin Samson Ayinla [1], Olukunle Oladipupo Amoo [2], Akoh Atadoga [3], Temitayo Oluwaseun Abrahams [4, *], Femi Osasona [5] and Oluwatoyin Ajoke Farayola [6]

[1] University of Law Business School, Manchester, United Kingdom.
[2] Department of Cybersecurity, University of Nebraska at Omaha, United States of America.
[3] Independent Researcher, San Francisco, USA.
[4] Independent Researcher, Adelaide, Australia.
[5] Scottish Water, UK.
[6] Financial Technology and Analytics Department, Naveen Jindal School of Management. Dallas, Texas, USA.

Abstract

In an era where artificial intelligence (AI) increasingly intersects with every facet of human life, the imperative for ethical AI has never been more pronounced. This paper delves into the complex interplay between technological advancements in AI and the overarching human values that guide societal norms. The background of the study establishes the urgency of addressing ethical challenges inherent in AI, such as privacy, bias, and accountability, within the broader context of regulatory and policy frameworks.

Aiming to critically evaluate the integration and effectiveness of ethical principles in AI applications, the paper navigates through a qualitative analysis, employing theoretical frameworks to dissect the ethical dimensions of AI. The scope encompasses a diverse range of topics, including global trends in ethical AI development, the impact of AI on human rights and personal freedoms, and the analysis of bias and fairness in AI algorithms. Real-world case studies provide insights into the successes and failures of ethical AI implementation, while the role of public perception and trust in AI adoption is scrutinized.

The main conclusions reveal a dynamic global landscape of ethical AI, emphasizing the need for robust ethical frameworks and proactive strategies to mitigate biases and ensure equitable outcomes. Recommendations advocate for clear ethical guidelines, integration of ethics in AI development, transparency, accountability, multi-stakeholder collaboration, public engagement, and continuous ethical evaluation.

The study concludes that balancing technological innovation with ethical constraints is crucial for the responsible development of AI. It underscores the importance of ethical vigilance, ensuring AI aligns with societal values and individual rights.

Keywords: Ethical AI; Artificial Intelligence; Privacy; Bias in AI; Human Values

1. Introduction

1.1. Defining Ethical AI: Concepts and Context

The concept of Ethical AI encompasses a broad spectrum of principles and practices aimed at ensuring that artificial intelligence (AI) technologies are developed and used in a manner that is beneficial and non-harmful to society. This

---

* Corresponding author: Temitayo Oluwaseun Abrahams.

involves a careful balance between technological advancement and adherence to ethical standards that reflect human values and societal norms.

At the core of Ethical AI lies the principle of beneficence, which is the obligation to act for the benefit of others. In the context of AI, this principle is particularly relevant in healthcare, where AI-assisted technologies are increasingly being employed. Lysaght et al, (2019) conducted a study on AI-assisted decision-making in healthcare and the application of an ethics framework for big data in health and research. The study focused on the ethical issues that arise with the use of AI in healthcare, including accountability and transparency of decisions made by AI-based systems, potential group harms arising from algorithmic bias, and the professional roles and integrity of clinicians. The authors applied the deliberative balancing approach of the Ethics Framework for Big Data in Health and Research to demonstrate how decision-makers can develop and implement AI-assisted support systems into healthcare and clinical practice ethically and responsibly

A vital element of Ethical AI involves refraining from anthropomorphizing AI systems. Dignum (2019) warns against ascribing human-like traits such as 'ethics' to AI, which may lead to overestimated expectations of AI's inherent ethical capabilities. Instead, the emphasis should be on the ethical application of AI, covering its entire lifecycle from conception and design to its operational use and maintenance. This strategy involves training AI developers, leaders, and users in ethics and fundamental AI principles, ensuring AI model transparency, implementing human-centered design, and establishing best practices for human-AI interaction (Dignum, 2019).

The role of law and regulation in governing the responsible development and use of AI is also a crucial component of Ethical AI. Ford (2021) critiques the prevailing trend of relying primarily on ethics-based frameworks for AI governance. He argues for a more robust legal and regulatory grounding to ensure adherence to ethical principles and to provide consequences for non-compliance. This approach recognizes the limitations of ethical frameworks in effectively governing private and corporate AI activities and underscores the need for legal mechanisms to enforce ethical principles (Ford, 2021).

Defining Ethical AI involves understanding and applying a range of ethical principles in a context-specific manner, particularly in sensitive areas like healthcare. It requires a shift from viewing AI as inherently ethical to focusing on the ethical use of AI throughout its lifecycle. Additionally, it calls for a balanced approach that includes both ethical frameworks and legal regulations to ensure responsible AI development and use. This approach aligns with the broader objective of balancing technological advancements with human values, ensuring that AI serves the greater good while respecting individual rights and societal norms.

1.2. Historical Overview of AI Development and Ethical Considerations

The historical evolution of artificial intelligence (AI) and its ethical considerations is a complex and multifaceted narrative that reflects the rapid advancements in technology and the corresponding emergence of ethical challenges. This journey from the inception of AI to the present day reveals a dynamic interplay between technological innovation and the evolving understanding of ethical principles in the context of AI.

In the early stages of AI development, ethical considerations were not a primary focus. However, as AI technologies began to demonstrate their potential to transform human-society-device relationships, the need to address ethical, legal, and standardization questions became increasingly urgent. Mezgár (2021) provides an overview of this transition, highlighting the shift from general ethical issues to the development of AI standards, particularly in the context of Cyber-Physical Production Systems (CPPS). This shift marks a critical point in AI history, where the focus expanded from technological development to include the ethical implications of these technologies (Mezgár, 2021).

The field of AI ethics, as a response to these growing concerns, has emerged as a subset of digital ethics. Kazim and Koshiyama (2021) offer a high-level conceptual discussion of AI ethics, tracing its roots to predecessors such as engineering ethics, philosophy of technology, and science and technology studies. They explore the evolution of AI ethics through various approaches, including principles, processes, and ethical consciousness, and discuss the central themes in translating ethics into engineering practice. This historical perspective underscores the nascent nature of AI ethics and its critical role in shaping the development and deployment of AI technologies (Kazim & Koshiyama, 2021).

The ethical debate around AI gained mainstream attention with the advent of Neural Networks and Machine Learning techniques, which significantly enhanced AI's utility and societal impact. Hagendorff (2020) evaluates the ethics of AI ethics by analyzing various guidelines released in recent years. These guidelines, comprising normative principles and recommendations, aim to harness the disruptive potentials of new AI technologies. Hagendorff's work provides a

comprehensive overview of the field of AI ethics and examines the extent to which ethical principles and values are implemented in the research, development, and application of AI systems (Hagendorff, 2020).

The historical overview of AI development and its ethical considerations reveals a trajectory marked by rapid technological advancements and a growing awareness of the ethical implications of these technologies. From the initial focus on technological innovation, the field has evolved to include a comprehensive framework of ethical guidelines and standards. This evolution reflects the ongoing effort to balance the transformative potential of AI with the need to uphold ethical principles and ensure the responsible development and use of AI technologies. The journey of AI ethics is characterized by its interdisciplinary nature, drawing from various fields to address the complex ethical challenges posed by AI in contemporary society.

1.3. The Intersection of Technology and Human Values in AI

The intersection of technology and human values in the realm of artificial intelligence (AI) is a critical area of study, as it encapsulates the challenges and opportunities presented by the integration of advanced technologies into the fabric of human society. This intersection is characterized by a dynamic tension between the capabilities of AI systems and the ethical, moral, and societal values that guide human behavior.

Stark and Hoey (2021) delve into the ethical considerations surrounding the integration of emotions in artificial intelligence systems. Their work highlights the complexities inherent in aligning AI with human values, particularly in the context of emotional responses and interactions. The authors discuss various definitions formulated by the AI community, emphasizing the importance of critically considering the consequences of adopting specific formulations of these values. Stark and Hoey's analysis provides a nuanced perspective on the ethical implications of developing AI systems capable of understanding and emulating human emotions.

Umbrello (2021) discusses the role of engineers in harmonizing human values with AI system design. He posits that most engineering innovations, driven by economic concerns, often overlook the societal values they impact. Umbrello argues for a system thinking approach in AI design, enabling engineers to understand the sociotechnical intricacies of AI systems across different sociocultural domains. He suggests that value-sensitive design, particularly through the use of envisioning cards, can help designers align human values with economic values, ensuring that the former are not sacrificed for the latter (Umbrello, 2021).

In the healthcare domain, the incorporation of AI brings forth distinct challenges and opportunities at the nexus of technology and human ethics. Jorstad (2020) delves into the profound effects of AI in healthcare, especially in the realms of medical diagnosis and treatment planning. Recognizing AI's proficiency in deciphering intricate medical data, Jorstad highlights the critical need to maintain elements of empathy and intuition in medical practices. The article argues that AI's role should be to augment, not replace, physicians, calling for an update in medical education to encompass technical and data analysis skills. This viewpoint emphasizes the importance of a balanced integration of technological innovation and human insight in healthcare, aiming to enhance patient care while preserving fundamental medical values (Jorstad, 2020).

The intersection of technology and human values in AI is a multifaceted domain that requires careful consideration of ethical, moral, and societal implications. The integration of AI into various sectors, including healthcare, highlights the need for a balanced approach that respects human values while leveraging the capabilities of AI. This balance is crucial for ensuring that AI technologies are developed and used in a manner that is beneficial and non-harmful to society, reflecting the broader objective of aligning technological advancements with human values.

1.4. Key Ethical Challenges in AI: Privacy, Bias, and Accountability

he rapid advancement of artificial intelligence (AI) has brought to the forefront a range of ethical challenges, particularly in the areas of privacy, bias, and accountability. These challenges are critical to address as they directly impact the trustworthiness and societal acceptance of AI technologies.

Jameel and Toheed (2020) critically examine the ethical dimensions of artificial intelligence, particularly focusing on the research challenges and potential solutions. Their study delves into the implications of AI in surveillance, exploring its potentials while also considering the ethical concerns related to data mining and personalization. The paper emphasizes the need for a balanced approach in AI development, ensuring that its benefits do not compromise individual privacy and ethical standards. This research contributes to the ongoing discourse on AI ethics, offering insights into how AI can be developed and utilized responsibly.

Bias in AI systems is another critical ethical challenge that Khan and Fadziso (2020) discuss in their article on the ethical issues surrounding the utilization of AI, robotics, and automation technologies. The authors argue that addressing bias in AI systems and algorithms is essential to ensure fairness and equity in decision-making processes. They emphasize the need for recognizing and mitigating biases in AI systems and the importance of collaboration among stakeholders to achieve this goal. The authors recommend an interdisciplinary approach to ethical design and the inclusion of ethics in the skill set of AI developers.

Accountability in AI refers to the ability to trace and hold responsible the decisions made by AI systems. Pasricha (2022) explores the ethical challenges in integrating AI into smart healthcare products, emphasizing the need for transparency, safety, responsibility, justice, and autonomy. The article outlines the open challenges and recommendations for integrating ethical principles into the design, validation, clinical trials, deployment, monitoring, repair, and retirement of AI-based smart healthcare products. Ensuring accountability in AI systems is crucial for maintaining public trust and for the responsible deployment of AI in sensitive areas like healthcare (Pasricha, 2022).

The key ethical challenges in AI of privacy, bias, and accountability are interconnected and require a multifaceted approach to address. Privacy concerns necessitate stringent data protection measures and respect for individual rights. Bias in AI requires a proactive approach to identify and mitigate discriminatory patterns in AI algorithms. Accountability involves establishing clear guidelines and frameworks for AI decision-making processes, ensuring that AI systems are transparent, safe, and aligned with ethical principles. Addressing these challenges is essential for the responsible development and deployment of AI technologies, ensuring they are beneficial and non-harmful to society.

1.5. Regulatory and Policy Frameworks Guiding Ethical AI

The landscape of ethical AI is increasingly being shaped by regulatory and policy frameworks, which are essential in ensuring that AI development aligns with societal values and human rights. Cooreman and Zhu (2022) delve into the challenges faced by existing frameworks in the ethical regulation of AI. They identify key issues such as the complexities in defining AI, the necessity for public participation in democratic governance, and the environmental impacts of AI systems. To navigate these challenges, they propose a multidimensional approach for defining AI, advocating for a relational perspective that acknowledges the cultural context of AI systems. Additionally, they suggest integrating environmental studies in assessing AI, aiming to foster environmentally sustainable AI technologies that can address global challenges (Cooreman & Zhu, 2022).

Madhavan et al. (2020) discuss the efforts in the U.S. Government to integrate AI into existing legal, ethical, and regulatory frameworks. Their work highlights the pivotal role of policy and governance in maximizing the benefits of AI technologies while mitigating risks and societal concerns. The involvement of the American Association for the Advancement of Science (AAAS) Science & Technology Policy Fellows in AI-related issues underscores the importance of developing trustworthy and responsible AI policies. This approach emphasizes the need for collaborative efforts among various stakeholders, including policymakers, researchers, and communities, to develop AI policies that are ethically grounded and socially beneficial (Madhavan et al., 2020).

These perspectives underscore the complexity of regulating AI and the importance of multi-dimensional approaches that consider cultural, environmental, and societal factors. The development of these frameworks is crucial for guiding AI towards ethical applications and ensuring that its advancements are in harmony with human values and rights.

1.6. Stakeholder Perspectives: Developers, Users, and Society

The landscape of ethical AI is shaped by a diverse array of stakeholders, each bringing unique perspectives and concerns. Understanding these perspectives is crucial for the development of AI systems that are not only technologically advanced but also ethically sound and socially responsible.

Deshpande and Sharp (2022) explore the wide range of stakeholders involved in responsible AI systems, including individual stakeholders like users, developers, and researchers, as well as organizational and national/international stakeholders. They emphasize the importance of identifying relevant guidelines for each stakeholder group to ensure that AI systems are built responsibly. This approach helps in addressing the specific needs and concerns of different stakeholders, thereby aligning AI development with broader societal values (Deshpande & Sharp, 2022).

Siqueira de Cerqueira, Acco Tives and Dias Canedo (2021) focus on the ethical guidelines and principles that are pertinent to AI development. They highlight the need for developers and product owners to be aware of these guidelines to ensure ethical considerations are integrated into the software development process. This work contributes to various stakeholders involved in AI development, including policymakers, ethicists, users, organizations, data scientists, and

development teams. By providing a body of knowledge in the field of AI ethics, this study aids stakeholders in developing systems that are ethically sound and aligned with societal expectations (Siqueira de Cerqueira, Acco Tives & Dias Canedo, 2021).

The perspectives of different stakeholders in ethical AI highlight the multifaceted nature of AI development. Developers, users, and society at large each have distinct concerns and expectations from AI systems. Addressing these diverse perspectives requires a comprehensive approach that incorporates ethical guidelines and principles into the AI development process. This approach ensures that AI systems are not only technologically robust but also ethically responsible and socially beneficial.

1.6.1. Case Examples of Ethical Dilemmas in AI Applications

The application of artificial intelligence (AI) across various sectors has surfaced numerous ethical dilemmas, challenging traditional norms and necessitating a re-evaluation of ethical frameworks. Vousinas et al. (2022) provide a comprehensive overview of these dilemmas, identifying key issues such as privacy concerns, algorithmic bias, and the impact of AI on employment. The study underscores the importance of adopting best practices to navigate these challenges, advocating for a holistic approach to address the ethical issues arising from AI systems

In the realm of healthcare, the insightful study by Alzahrani et al. (2021) shed light on the critical usability attributes of AI-enabled eHealth mobile applications. Central to their research is the integration of ethical principles throughout the AI development lifecycle, highlighting the essential role of ethical considerations. The study is particularly notable for identifying seven ethical principles that are crucial in the development and deployment of these applications: autonomy, beneficence, non-maleficence, justice, privacy, transparency, and accountability. These principles serve as a foundational guide for developers and stakeholders, ensuring the creation of eHealth solutions that are not only user-friendly but also ethically responsible

Cobianchi et al. (2022) explore the ethical dilemmas in applying AI to surgery, highlighting issues related to human agency, accountability for errors, technical robustness, privacy, data governance, transparency, diversity, non-discrimination, and fairness. The study suggests that addressing these ethical issues may require expanding the focus of surgical AI research to include implementation science and digital health education, balancing the opportunities offered by emerging AI technologies with respect for ethical principles in patient-centric philosophy.

Stahl and Stahl (2021) review the ethical benefits and challenges of AI development, deployment, and use. The study categorizes ethical issues arising from different AI technologies, including machine learning, artificial general intelligence, and socio-technical systems incorporating AI. It provides detailed accounts of ethical issues such as the impact on society, decision-making processes, and the potential for AI to exacerbate inequalities.

These case studies collectively illustrate the diverse and complex nature of ethical dilemmas in AI applications. From general applications to specific fields like healthcare and surgery, these dilemmas require careful consideration and a tailored approach. Addressing these issues involves not only understanding the specific ethical principles relevant to each domain but also integrating these principles into the AI development process. This approach ensures that AI systems are developed responsibly, with a focus on ethical considerations and societal impact.

1.6.2. The Role of Ethical AI in Sustainable Technological Growth

The integration of ethical considerations in the development and application of artificial intelligence (AI) is crucial for ensuring sustainable technological growth. Mantini (2022) introduces the concept of 'Algor-ethics', which emphasizes the need for a dedicated study on the ethics applied to technology, algorithms, and AI. The paper proposes an application of the concept of sustainability to technological processes, particularly in the development of AI systems. This approach termed the Dynamical Techno-Algor-Ethical Composition, aims to define the interaction between ethical ingredients involving the human person in relation to technology, taking a person-centered approach. The framework seeks to establish a structure and definition of Technological Sustainability, evaluating the process of AI algorithm development as a concrete application of the analyzed framework.

Maharani, Amin and Taufiqoh (2022) analyze the development of AI in Indonesia, offering an ethical paradigm based on three principles: responsibility, awareness, and sustainability. This paradigm serves as a basis for the design of AI vision, strategy, and implementation, managing the impacts of AI development, including its positive potential. The research highlights the symbiotic relationship between human-technology and human-nature, emphasizing the role of ethics in maintaining humanity and stability of life. The ethical paradigm proposed is intended to guide AI development in a way that anticipates and prevents technology from becoming a disaster in the future.

The role of ethical AI in sustainable technological growth involves integrating ethical principles and frameworks in the development and application of AI technologies. This integration ensures that AI development is aligned with human values and societal needs, promoting a sustainable and responsible technological future. The concept of Algor-ethics and the ethical paradigms proposed in these studies provide a foundation for developing AI systems that are not only technologically advanced but also ethically sound and beneficial to society.

1.7. Research Gap: Evaluating the Efficacy of Ethical Guidelines and Principles in Real-World AI Applications

The efficacy of ethical guidelines and principles in real-world AI applications is a critical area of research, as it determines how well these theoretical frameworks translate into practical outcomes. Wei and Zhou (2022) delve into this issue by conducting a content analysis of the AI Incident Database, which catalog real-world AI failures. Their study identifies 13 application areas where unethical use of AI is prevalent, including intelligent service robots, language/vision models, and autonomous driving. The research categorizes ethical issues into eight forms, ranging from inappropriate use and racial discrimination to physical safety and unfair algorithms. This taxonomy aims to provide AI practitioners with practical guidelines for deploying AI applications ethically.

Leimanis (2020) evaluates self-imposed ethical guidelines for AI in education, highlighting the global discourse on ethical principles and values in AI applications to education. The paper analyzes various guidelines and policies published by research institutions, companies, public agencies, and non-governmental entities, noting significant differences in content and application. The study aims to facilitate further discussion about ethical principles and responsibilities of educational institutions using AI, seeking to arrive at a consensus concerning safe and desirable uses of AI in education.

Cerqueira (2021) address the gap between AI ethics principles and practical guidance for developers. They present the RE4AI Ethical Guide, evaluated through a focus group with AI professionals. The guide comprises 26 cards across 11 principles, aiming to facilitate the elicitation of ethical requirements in AI development. The preliminary results reveal the guide's potential to assist developers in operationalizing ethics in AI, contributing to bridging the gap between principles and practice.

Evaluating the efficacy of ethical guidelines and principles in real-world AI applications is essential for ensuring that AI technologies are developed and used responsibly. The research highlights the need for practical, actionable guidelines that can be effectively implemented in various AI application areas. By bridging the gap between theoretical ethical frameworks and practical implementation, these studies contribute to the development of AI technologies that are not only technologically advanced but also ethically sound and socially responsible.

1.8. Objectives and Scope of the Research

The primary aim of this research is to critically examine and evaluate the integration and effectiveness of ethical principles and guidelines in the development and application of artificial intelligence (AI) technologies. This study focuses on understanding how ethical AI can be operationalized in real-world scenarios, ensuring that AI development aligns with human values and societal norms.

A key objective is to assess current ethical AI frameworks, analyzing existing guidelines and frameworks for AI. This involves a comprehensive review of the principles laid out in various ethical AI guidelines and evaluating their applicability and effectiveness across different AI application areas. The research will delve into the identification and categorization of key ethical challenges faced in AI applications, such as issues of privacy, bias, and accountability. This includes examining case studies and real-world incidents where AI has posed ethical dilemmas, providing a nuanced understanding of the ethical landscape in AI.

Another crucial objective is to evaluate the efficacy of ethical AI in practice. This includes assessing the practical implementation of ethical guidelines and principles in real-world AI applications and determining whether these guidelines have a tangible impact on the development and deployment of AI technologies. The study will explore the perspectives of various stakeholders involved in AI development and application, including developers, users, and society at large, to understand how these stakeholders perceive ethical AI and the impact of AI technologies on their interests and concerns.

In summary, the research aims to bridge the gap between theoretical ethical AI frameworks and their practical application. By focusing on the assessment of current frameworks, the identification of ethical challenges, and the evaluation of the efficacy of ethical AI in practice, the study contributes to the responsible and sustainable development

of AI technologies. The scope of the research encompasses a broad range of AI applications, stakeholder perspectives, and ethical challenges, aiming to provide comprehensive insights into the ethical dimensions of AI.

2. Methodology

2.1. Research Approach: Qualitative Analysis and Theoretical Frameworks in AI Ethics

The research approach for this study is anchored in qualitative analysis and the application of theoretical frameworks in AI ethics. This methodology is essential for comprehending the intricate ethical dimensions of AI, which often transcend quantifiable metrics and delve into the realms of moral philosophy, societal norms, and human values.

Taddeo, McCutcheon and Floridi (2019) critically examine the complexities of trusting artificial intelligence (AI) in the realm of cybersecurity, describing it as a double-edged sword. Their analysis underscores the importance of a risk-based approach in ethical AI assessments, emphasizing the need for qualitative risk analysis. The paper delves into the ethical implications of AI innovations in cybersecurity, highlighting the balance between ethical considerations and technical challenges. Taddeo, McCutcheon and Floridi (2019) advocate for a nuanced understanding of AI's role in cybersecurity, acknowledging both its potential benefits and inherent risks.

Franzke (2022) contributes to this discourse by conducting a qualitative analysis of international ethics guidelines in AI. The study highlights the diversity of values promoted in these guidelines and underscores the need for greater reflexivity in their ethical frameworks. This analysis provides valuable insights into the ethical discourse surrounding AI, offering a nuanced understanding of how ethical principles are operationalized in various AI contexts (Franzke, 2022).

Georgieva et al. (2022) further this exploration by mapping AI ethical principles onto the lifecycle of AI-based digital services and products. Their work addresses the operationalization of ethics in AI, a critical aspect often overlooked in traditional ethical frameworks. By combining AI ethics with explicit governance models, they shed light on the practical, conceptual, and political implications of ethical AI, highlighting key challenges in operationalizing AI ethics (Georgieva et al., 2022).

The qualitative approach adopted in this study, supported by the theoretical frameworks discussed, provides a comprehensive understanding of the ethical landscape in AI. This approach not only facilitates a deeper understanding of the ethical challenges in AI but also aids in developing practical strategies for integrating ethical considerations into AI development and deployment. The insights gained from this research will contribute significantly to the field of AI ethics, offering guidelines for ethical AI practices that are both theoretically sound and practically viable.

2.2. Criteria for Selecting Case Studies and Academic Sources

The need for comprehensive and diverse perspectives guides the criteria for selecting case studies and academic sources in AI ethics. Ryan et al. (2021) emphasize the importance of juxtaposing academic discourse with organizational reality, suggesting that case studies should be chosen based on their ability to provide real-world insights into the ethical application of AI technologies. This approach ensures that theoretical discussions are grounded in practical experiences (Ryan et al., 2021).

Alzahrani et al. (2021) stresses the critical role of context in the realm of AI ethics, advocating for in-depth case studies that demonstrate the application of ethical principles in specific sectors, such as healthcare. Their study highlights the need for these case studies to not only expose ethical challenges but also to offer actionable solutions and frameworks for addressing these challenges in practical environments (Alzahrani et al., 2021)

Alzahrani et al. (2021) highlights the significance of context in AI ethics, advocating for case studies that demonstrate the operationalization of ethical principles in specific domains, such as healthcare. Their work underscores the need for case studies that not only illustrate ethical challenges but also offer solutions and frameworks for addressing these challenges in practical settings (Alzahrani et al., 2021).

Birhane et al. (2022) argue for the inclusion of marginalized perspectives in AI ethics research. They suggest that academic sources should be selected based on their focus on underrepresented groups and their ability to address power asymmetries in AI development and application. This criterion ensures that the ethical analysis is inclusive and considers the impacts of AI on all segments of society (Birhane et al., 2022).

The selection of case studies and academic sources for this research is based on their relevance to real-world applications, their focus on specific contexts and operational challenges in AI ethics, and their inclusion of diverse and marginalized perspectives. This approach ensures a holistic understanding of AI ethics, encompassing both theoretical and practical dimensions.

3. Results

3.1. Trends in Ethical AI Development: Global Perspectives

The development of ethical AI is a global endeavor, with varying approaches and emphases across different regions. Daly et al. (2020) provide an insightful overview of AI governance and ethics from a global perspective, highlighting the initiatives and developments in countries like Australia, China, the European Union, India, and the United States. Their analysis reveals that while AI governance and ethics initiatives are most advanced in China and the EU, the US has made significant progress in recent years. India, however, remains an outlier in not articulating a set of AI ethics principles, and Australia faces challenges as a smaller player in forging its path in AI ethics. This study underscores the shift in focus from high-level principles to legally enforceable outcomes in AI ethics (Daly et al., 2020).

A study by Daly et al. (2019) delved into the global conversation on AI ethics, examining the ethical standards and governance initiatives issued by various actors in different countries and sectors. They note the increasing prevalence of AI in society and the economy, accompanied by concerns over problematic implementations, particularly in sensitive areas like the military, medicine, and criminal justice. The study emphasizes the need for AI systems to adhere to ethical standards and highlights the global efforts to address these concerns (Daly et al., 2019).

Crockett, Colyer and Latham (2021) explore the ethical landscape of data and AI from the perspective of citizens. Their research, based on a longitudinal survey, captures citizen perspectives on the ethical dimensions of AI. The study finds that perceptions of AI, trust, bias, and fairness vary depending on specific AI applications and contexts. It also reveals a strong interest among citizens in undertaking educational courses on AI concepts, indicating a growing public desire to engage with and understand AI technologies. This citizen-centric approach to AI ethics highlights the importance of public involvement and education in shaping ethical AI development (Crockett, Colyer & Latham, 2021).

The trends in ethical AI development reflect a diverse global landscape with varying degrees of advancement and focus. While some regions are leading in establishing enforceable ethical standards for AI, others are still developing their approaches. The involvement of citizens in understanding and shaping the ethical use of AI emerges as a crucial element in this global endeavor. These studies collectively provide a comprehensive view of the current state and future directions of ethical AI development from a global perspective.

3.2. Impact of AI on Human Rights and Personal Freedoms

The impact of Artificial Intelligence (AI) on human rights and personal freedoms is a multifaceted issue that encompasses various aspects of modern life. Lubin's work delves into the implications of AI in military contexts, particularly focusing on the rights to privacy and data protection in wartime AI applications. The study highlights the challenges posed by the involvement of private technology contractors in military projects, which often bring large sets of personal data and commercial practices that may conflict with international human rights and civil liberties. This research underscores the need for robust regulation of AI development and deployment in military settings to protect digital freedoms and civil liberties (Lubin, 2022).

Rodrigues' research addresses the broader societal impacts of AI, critiquing the dominant vision of AI as autonomous systems outperforming humans. The article discusses the misconstrued vision of AI and alternative visions of AI that focus on augmenting human creativity and supporting successful digital technologies. The study highlights the legal and human rights issues of AI, including algorithmic transparency, cybersecurity vulnerabilities, unfairness, bias and discrimination, lack of contestability, legal personhood issues, intellectual property issues, adverse effects on workers, privacy and data protection issues, liability for damage, and lack of accountability. The article uses the frame of vulnerability to consolidate the understanding of critical areas of concern and guide risk and impact mitigation efforts to protect human well-being (Rodrigues, 2020).

Adams and Loideáin (2019) explore the gendered implications of AI, particularly in the context of virtual personal assistants like Siri and Alexa. Their study critiques the reproduction of harmful gender stereotypes in these AI systems, which often portray female characters in subordinate roles. The research assesses how this design constitutes indirect discrimination under international women's rights law and explores potential legal responses to address these issues.

This work highlights the importance of considering gender perspectives in AI development to prevent the reinforcement of negative stereotypes and ensure the protection of women's rights (Adams & Loideáin, 2019).

The impact of AI on human rights and personal freedoms is a complex issue that requires careful consideration of various factors, including privacy, data protection, societal impacts, and gender perspectives. The studies discussed provide valuable insights into the challenges and opportunities presented by AI in protecting and promoting human rights and personal freedoms. They underscore the need for ethical AI development that is mindful of its broader societal implications and adheres to international human rights standards.

3.3. Analysis of Bias and Fairness in AI Algorithms

The analysis of bias and fairness in AI algorithms is a critical aspect of ethical AI development. Zhou et al. (2021) provide an overview of the challenges associated with bias and fairness in machine learning (ML) algorithms. They discuss the types and sources of data bias and the nature of algorithmic unfairness. Their work includes a review of fairness metrics in the literature and a discussion of their limitations. The authors also describe de-biasing techniques in the model life cycle, emphasizing the need for proactive measures to mitigate bias and ensure fairness in AI applications (Zhou et al., 2021).

In a related study, Zhou et al. (2022) delve deeper into the issues of bias and fairness in AI and ML algorithms. They begin with a discussion of bias and fairness issues that arise with the use of AI techniques, focusing on supervised machine learning algorithms. The study describes the types and sources of data bias and discusses the nature of algorithmic unfairness. Additionally, the authors provide a review of fairness metrics in the literature, discuss their limitations, and describe de-biasing techniques in the model life cycle. This comprehensive overview highlights the importance of addressing bias and fairness in AI development (Zhou et al., 2022).

Puyol-Antón et al. (2021) investigate the fairness of deep learning models in cardiac MR image analysis, focusing on the problem of training data imbalance. They perform an analysis for racial/gender groups and find statistically significant differences in performance between different racial groups. The study explores strategies to reduce racial bias, including stratified batch sampling, fair meta-learning for segmentation, and protected group models. The results demonstrate that the racial bias results from the use of imbalanced training data, and that all proposed bias mitigation strategies improved fairness. This research underscores the importance of considering demographic factors in AI development to ensure equitable outcomes (Puyol-Antón et al., 2021).

The analysis of bias and fairness in AI algorithms is a complex and multifaceted issue that requires careful consideration of data sources, algorithmic design, and the intended use of AI systems. The studies discussed provide valuable insights into the challenges of ensuring fairness in AI and offer practical strategies for mitigating bias. These efforts are crucial for the development of AI technologies that are not only technologically advanced but also ethically sound and socially responsible.

3.4. Case Studies: Successes and Failures in Ethical AI Implementation

The implementation of ethical AI practices has seen a mix of successes and failures, as evidenced by various case studies. Ryan et al. (2021) present a comprehensive analysis of ethical issues in AI through a multi-case study approach. Their research juxtaposes academic discourse with organizational reality, offering insights into the ethical challenges faced by organizations in different domains. The study categorizes ethical issues identified in the literature and compares them with real-world cases, highlighting the overlap and correlation between theoretical concerns and practical challenges. This approach provides a nuanced understanding of the ethical landscape in AI and the need for targeted solutions to address these issues effectively.

Park and Bae (2022) explore the ethical implications of AI through specific case studies. Their research examines events that occur in the 'uncanny valley', where AI systems exhibit near-human characteristics, leading to complex ethical dilemmas. The study also addresses the role of AI as a major agent in new virtual and offline spaces, raising concerns about ownership disputes and the ethical use of AI in capitalist markets. The authors conclude that AI's ability to change its algorithms autonomously presents unique challenges, necessitating preparation for the next steps in AI development.

Levantino (2020) on the other hand conducted a human rights-based study on AI, focusing on the case of Facial Recognition Technology (FRT). The study investigated, analyzed, and discussed several threats FRT can pose to human rights, democracy, and the rule of law. The author framed the uses of FRT by law enforcement authorities adopting the European human rights law framework. The study unveiled that the risks connected to the deployment of FRT are

increased when advocated for the pursuit of "public security". Based on the performed analysis, it can be concluded that, while proper regulations would mitigate the adverse effects generated by FRT, the general public should be more sensitive to data protection and privacy issues to enable an environment for "human flourishing".

These case studies highlight the need for a deeper understanding of ethical challenges in AI, the importance of aligning theoretical principles with practical realities, and the necessity for organizations to develop effective strategies for managing ethical crises in AI technologies. These studies contribute to the ongoing discourse on ethical AI, offering lessons learned and best practices for future AI development and deployment.

3.4.1. Role of Public Perception and Trust in AI Adoption

The role of public perception and trust in AI adoption is a critical factor influencing the integration and acceptance of AI technologies in society. Khan's research emphasizes the importance of understanding public perception as a determinant of trust in AI. The study proposes measuring the impact of public perception on trust in AI, focusing on two dimensions: control of AI and ethics in AI. These dimensions, along with a mediating factor called mood, are crucial for building trust in AI technologies (Khan, 2022).

Arrieta et al. (2020) delve into the realm of Explainable Artificial Intelligence (XAI), offering a comprehensive overview of its concepts, taxonomies, and the opportunities and challenges it presents for responsible AI. The study explores how XAI can shape public perceptions of AI, emphasizing the importance of personal evaluations in understanding AI systems. It also addresses the increasing concern of cybersecurity threats within AI, highlighting the need for transparent and explainable models. Furthermore, the influence of user diversity on AI evaluation is examined, suggesting that varied user backgrounds can significantly impact the interpretation and trust in AI systems. Arrieta et al. (2020) propose suggestions for advancing XAI, focusing on its role in fostering responsible and trustworthy AI development.

Yigitcanlar, Degirmenci and Inkinen (2022) investigate the drivers behind public perception of AI in major Australian cities. The study reveals concerns about privacy invasion and lower trust in companies and government deploying AI. It also finds appreciation for AI's benefits in urban services and disaster management. The research underscores the need for policies to minimize public concerns and maximize AI awareness (Yigitcanlar, Degirmenci & Inkinen, 2022).

3.4.2. Ethical Considerations in Emerging AI Technologies

Emerging AI technologies bring forth new ethical considerations that need to be integrated into their development and deployment. Antoniou's work focuses on the necessity of teaching and learning ethics for AI, especially in the context of emerging technologies. The chapter emphasizes that AI developers need to understand ethics as part of their skill set to ensure ethical design in AI products (Antoniou, 2021).

Subramanian (2017) addresses the legal, security, privacy, and ethical issues associated with emergent AI systems, particularly social robots. The paper explores the history of the field and delves into legal and ethical considerations, proposing solutions to these challenges. The study emphasizes the need for future research to address the challenges posed by AI-enabled systems (Subramanian, 2017).

Kim and Hong (2022) examine ethical considerations in AI-based technologies in healthcare, particularly in nursing. The study investigates various types of AI technologies applied to nursing and explores ethical guidelines and issues based on biomedical ethics and the nursing code of ethics. The research highlights the importance of considering patient safety and non-discrimination in AI applications in healthcare (Kim & Hong, 2022).

The role of public perception and trust in AI adoption is crucial for the successful integration of AI technologies in society. Ethical considerations in emerging AI technologies are equally important, requiring a multidisciplinary approach that includes legal, security, privacy, and ethical aspects. These studies provide valuable insights into the factors influencing public perception and trust in AI and the ethical challenges posed by emerging AI technologies.

4. Analysis of findings

4.1. Balancing Innovation with Ethical Constraints in AI

The balance between innovation and ethical constraints in AI is a critical aspect of responsible AI development. This is the main theme of Rayhan and Rayhan's (2021) article, which examines the complex relationship between AI and human rights. The authors review the historical evolution of AI, its ethical challenges, and the legal frameworks that regulate

its use. They also present case studies that illustrate how AI can both promote and violate human rights. The authors emphasize the need for ethically driven development and responsible deployment, explainable and transparent AI, and interdisciplinary collaboration to ensure ethical design. They argue that AI should be aligned with human values and respect individual liberties in the digital age.

In their article, Panait, Ljubenkov and Alic (2021) discuss the balance between innovation and regulation in AI and whether Europe is leading or lagging behind in this regard. The authors argue that AI is an emerging technology area that permeates political, economic, socio-cultural, and other spheres of life. They present a comparative analysis of the European Union, the United States of America, and China as geopolitical global players with vastly different AI leadership stances when it comes to national strategies, research & development funding, data protection acts, or ethical concerns. The authors recommend policy recommendations, either regulatory, strategic, or focused on public governance and technology management. They emphasize the importance of ethical considerations in AI, teaching and learning ethics as part of the skill set for AI developers, and an interdisciplinary approach to ethical design.

Rakova et al. (2021) delve into the practical aspects of implementing responsible AI in organizations, focusing on the perspectives of practitioners. Their study underscores the need for ethical AI systems that not only safeguard individual privacy but also adhere to ethical standards. They explore innovative algorithmic techniques and emphasize the importance of a comprehensive approach in shifting organizational practices towards responsible AI. The conclusion drawn from the study highlights the critical role of integrating ethical considerations into the fabric of AI development and deployment. This research offers valuable insights into the real-world challenges and enablers of adopting ethical AI practices in organizations.

Wyeld et al. (2007) explore the ethical dimensions of using the Torque Game Engine as a tool to preserve and promote Australian Aboriginal cultural heritage, particularly through Indigenous storytelling. Their work focuses on the application of AI in supporting the documentation and revitalization of Indigenous languages. The study carefully considers the balance between social impact and technical opportunities, alongside ethical constraints. It involves surveys to understand the community's perspective and emphasizes the importance of creating and deploying language technology ethically. This research highlights the potential of AI in safeguarding intangible cultural heritage while navigating the complex ethical landscape associated with such endeavors.

Balancing innovation with ethical constraints in AI requires a multidisciplinary approach that includes ethical considerations in the development process. The studies discussed highlight the importance of integrating ethical principles into AI design, ensuring privacy protection, and addressing the unique challenges posed by emerging AI technologies. These efforts contribute to the development of AI technologies that are not only technologically advanced but also ethically sound and socially responsible.

4.2. Evaluating the Effectiveness of Current Ethical Frameworks in AI

The effectiveness of current ethical frameworks in AI is a topic of significant importance, given the rapid advancement and integration of AI technologies in various sectors. Atkins, Badrie and van Otterloo (2021) conducted a practical evaluation of four commercial chatbots against four responsible AI frameworks. Their findings revealed that different ethical frameworks produced varying assessment scores, indicating that many ethical AI frameworks contain principles that are challenging to evaluate for anyone except the chatbot developer. The study also found that domain-specific ethical AI guidelines are more practical and easier to use than domain-independent frameworks, suggesting a need for more measurable standards in ethical AI guidelines (Atkins, Badrie & van Otterloo, 2021).

Erdogan, Altunel and Tarhan (2022) conducted a systematic analysis of tools and frameworks that support the development process of AI-based software. The study focused on ethical issues such as explicability, fairness, privacy, and accountability, as well as software and algorithms, frameworks, guidelines, or process models. The authors reviewed 14 primary studies and evaluated their proposals with respect to basic characteristics including theme, research methods, types, domains, and a number of cases in empirical validations. The review revealed that the domains of empirical validation are diverse while the number of empirical cases applied for validation is limited. The authors also observed that only half of the primary studies provide links to their proposals as open-source, which is very important for the repeatability of the empirical validations.

Peters et al. (2020) describe two complementary frameworks for integrating ethical analysis into engineering practice to help address the challenge of moving from principles to practice in ethical AI. They provide outcomes of an ethical analysis conducted within the specific context of Internet-delivered therapy in digital mental health. The study highlights the importance of these frameworks in providing tools and insights for data-enabled and intelligent

technology development more broadly, emphasizing the need for responsible AI that promotes transparency, minimizes risks of bias or error, and protects patient well-being (Peters et al., 2020).

The evaluation of current ethical frameworks in AI reveals a need for more practical and measurable standards, a focus on specific ethical issues, and the importance of integrating ethical considerations into the AI development process. These studies contribute to the understanding of how ethical principles can be effectively implemented in AI technologies, ensuring that they are developed and used responsibly and beneficially.

4.3. Societal and Cultural Influences on AI Ethics

The societal and cultural influences on AI ethics are profound and multifaceted, impacting the development, deployment, and perception of AI technologies across different regions and communities.

Hardman (2022) explores the cultural influences on AI, particularly in the context of China's investment in AI and its national strategy. The study highlights the geopolitical role of business and academic players in AI development and the societal implications of AI applications. It emphasizes the need to consider societal impacts alongside technological innovations, as AI applications have far-reaching consequences that extend beyond mere technological advancements. This research underscores the importance of understanding the cultural context in which AI is developed and deployed (Hardman, 2022).

Sanusi and Olaleye (2022) investigate the role of cultural competence and ethics in AI education, particularly in the K-12 sector. Their study, conducted in Nigeria, reveals that cultural competence and ethics significantly influence AI content. The research highlights the importance of integrating cultural elements and ethical considerations in AI education to ensure effective learning. The findings suggest that emphasizing cultural elements and humanistic thinking, along with ethical considerations, is crucial in designing AI content and instructional materials. This approach ensures that AI education is inclusive and considers the diverse cultural backgrounds of students (Sanusi & Olaleye, 2022).

Kassens-Noor, Siegel and Decaminada (2021) discuss the choice between ethics and morals as a determinant in embracing AI in future urban mobility. Their study examines respondents' preferences on whether autonomous vehicles' AI should learn their owners' personal morals or adopt societal ethics. The findings reveal a nearly even split between these preferences, highlighting the deep divide in individual choices regarding AI governance. This research illustrates the importance of considering societal ethics and individual morals in AI development, especially in applications that directly impact daily life, such as urban mobility (Kassens-Noor, Siegel & Decaminada, 2021).

Societal and cultural influences play a critical role in shaping AI ethics. These influences affect how AI technologies are perceived, developed, and utilized in different contexts. Understanding these influences is essential for developing AI technologies that are not only technologically advanced but also ethically sound and culturally sensitive. The studies discussed provide valuable insights into the complex interplay between societal values, cultural norms, and ethical considerations in AI.

4.4. Recommendations for Enhancing Ethical Practices in AI

Enhancing ethical practices in AI is essential for ensuring responsible development and deployment of AI technologies. Steinke, LaBrie and Sarkar (2022) emphasize the importance of continuous ethical analysis of AI algorithms. They propose a framework for ongoing ethical evaluation as AI applications are updated and enhanced, highlighting the necessity for a dynamic approach to ethics in AI. This approach ensures that ethical considerations keep pace with technological advancements and changing societal norms.

Liaw et al. (2020) provide practical recommendations for the ethical use of electronic health record data and AI in primary care. They stress the importance of consent and formal processes to govern access and sharing throughout the data lifecycle. Their recommendations also include attention to data governance and quality management, recognizing the need for new processes to address ethical issues arising from AI in healthcare. These recommendations are crucial for maintaining trust and ensuring the responsible use of AI in sensitive areas like healthcare.

Tonidandel, King and Cortina (2018) discuss the potential of big data methods to build organizational science. The authors argue that data science techniques such as data mining, data visualization, and machine learning can address numerous questions in the organizational sciences. They identify the biggest opportunities and obstacles afforded by big data and discuss how their methods will be most impacted by the data analytics movement. The authors recommend a list of resources to help interested readers incorporate big data methods into their existing research. They emphasize

the importance of advancing ethical data science and present an AI Ethics Maturity Model comprising six dimensions crucial for the operationalization of AI ethics within an organization.

4.5. Anticipating Future Ethical Challenges in AI Development

Anticipating future ethical challenges in AI development is crucial for preparing for the evolving landscape of AI technologies. Dara, Hazrati Fard and Kaur (2022) explore the ethical challenges in the use of AI in digital agriculture. They examine issues such as fairness, transparency, accountability, sustainability, privacy, and robustness in AI applications in farming. The study provides recommendations for agriculture technology providers (ATPs) and policymakers on mitigating ethical issues, covering a wide range of considerations like addressing farmers' privacy concerns, ensuring reliable AI performance, enhancing sustainability in AI systems, and reducing AI bias.

Enhancing ethical practices in AI and anticipating future ethical challenges require a comprehensive and dynamic approach. The recommendations and models discussed in these studies provide valuable guidance for integrating ethical considerations into AI development and deployment. These efforts contribute to the responsible advancement of AI technologies, ensuring they align with ethical standards and societal values.

5. Conclusion

This study embarked on a comprehensive exploration of ethical AI, meticulously addressing its aim and objectives to evaluate the integration and effectiveness of ethical principles in AI development and application. The research has successfully illuminated the multifaceted nature of ethical AI, providing a nuanced understanding of the challenges, opportunities, and implications of AI technologies in the context of human values and societal norms.

The study provides a dynamic global landscape of ethical AI development, with varying degrees of advancement and focus across regions. The research highlighted the critical impact of AI on human rights and personal freedoms, emphasizing the need for robust ethical frameworks to safeguard these fundamental aspects. The analysis of bias and fairness in AI algorithms underscored the necessity for proactive strategies to mitigate inherent biases and ensure equitable outcomes. Case studies on ethical AI implementation offered valuable insights into both the successes and failures in the field, serving as a practical guide for future endeavors in ethical AI development.

The role of public perception and trust in AI adoption emerged as a pivotal factor, influencing the integration and acceptance of AI technologies in society. Ethical considerations in emerging AI technologies were identified as crucial for future development, requiring a forward-looking approach that anticipates potential ethical challenges.

The study's recommendations for enhancing ethical practices in AI are grounded in the need for clear, actionable ethical guidelines, integration of ethics into the AI development process, promotion of transparency and accountability, fostering multi-stakeholder collaboration, and encouraging public engagement and education. Continuous ethical evaluation was highlighted as essential for ensuring that AI technologies remain aligned with ethical principles and societal values.

In conclusion, this study has met its objectives by providing a thorough analysis of ethical AI, offering a roadmap for balancing technological advancements with ethical constraints. The findings and recommendations presented serve as a foundation for future research and practice in ethical AI, contributing significantly to the responsible and sustainable development of AI technologies. The study underscores the imperative of ethical vigilance in AI development, ensuring that AI serves the greater good while respecting individual rights and societal norms.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Adams, R. and Loideáin, N.N., (2019). Addressing indirect discrimination and gender stereotypes in AI virtual personal assistants: the role of international human rights law. *Cambridge International Law Journal*, *8*(2), pp.241-257. DOI: 10.4337/CILJ.2019.02.04

[2] Alzahrani, A.S., Gay, V., Alturki, R. and AlGhamdi, M.J., (2021). Towards understanding the usability attributes of AI-enabled eHealth mobile applications. *Journal of Healthcare Engineering*, *2021*. https://doi.org/10.1155/2023/9896541

[3] Antoniou, J. and Antoniou, J., (2021). Dealing with emerging ai technologies: Teaching and learning ethics for AI. *Quality of Experience and Learning in Information Systems: Incorporating Learning and Ethics into Characterizations of Quality of Experience*, pp.79-93. DOI: 10.1007/978-3-030-52559-0_6

[4] Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R. and Chatila, R., (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, *58*, pp.82-115. https://doi.org/10.1016/j.inffus.2019.12.012

[5] Atkins, S., Badrie, I. and van Otterloo, S., (2021). Applying Ethical AI Frameworks in practice: Evaluating conversational AI chatbot solutions. *Computers and Society Research Journal*. DOI: 10.54822/qxom4114

[6] Birhane, A., Ruane, E., Laurent, T., S. Brown, M., Flowers, J., Ventresque, A. and L. Dancy, C., (2022). The forgotten margins of AI ethics. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 948-958). DOI: 10.1145/3531146.3533157

[7] Cerqueira, J.A.S.D., (2021). Exploring ethical requirements elicitation for applications in the context of AI. DOI: 10.5753/sbsi_estendido.2022.222269

[8] Cobianchi, L., Verde, J.M., Loftus, T.J., Piccolo, D., Dal Mas, F., Mascagni, P., Vazquez, A.G., Ansaloni, L., Marseglia, G.R., Massaro, M. and Gallix, B., (2022). Artificial intelligence and surgery: ethical dilemmas and open issues. *Journal of the American College of Surgeons*, *235*(2), pp.268-275. DOI: 10.1097/XCS.0000000000000242

[9] Cooreman, H. and Zhu, Q., (2022). Critical Reflections on the Ethical Regulation of AI: Challenges with Existing Frameworks and Alternative Regulation Approaches. In *2022 IEEE International Symposium on Technology and Society (ISTAS)* (Vol. 1, pp. 1-5). IEEE. DOI: 10.1109/ISTAS55053.2022.10227116

[10] Crockett, K., Colyer, E. and Latham, A., (2021). The ethical landscape of data and artificial intelligence: Citizen perspectives. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1-9). IEEE. DOI: 10.1109/SSCI50451.2021.9660153

[11] Daly, A., Hagendorff, T., Hui, L., Mann, M., Marda, V., Wagner, B., Wang, W. and Witteborn, S., (2019). Artificial intelligence governance and ethics: global perspectives. *arXiv preprint arXiv:1907.03848*.

[12] Daly, A., Hagendorff, T., Li, H., Mann, M., Marda, V., Wagner, B. and Wang, W.W., (2020). AI, governance and ethics: global perspectives. *University of Hong Kong Faculty of Law Research Paper*, (2020/051). DOI: 10.2139/ssrn.3684406

[13] Dara, R., Hazrati Fard, S.M. and Kaur, J., (2022). Recommendations for ethical and responsible use of artificial intelligence in digital agriculture. *Frontiers in Artificial Intelligence*, *5*, p.884192. DOI: 10.3389/frai.2022.884192

[14] Deshpande, A. and Sharp, H., (2022). Responsible AI Systems: Who are the Stakeholders? In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 227-236). DOI: 10.1145/3514094.3534187

[15] Dignum, V., (2019). *Responsible artificial intelligence: how to develop and use AI in a responsible way* (Vol. 2156). Cham: Springer.

[16] Erdogan, T.G., Altunel, H. and Tarhan, A.K., (2022). Tools/frameworks that support development process of AI-based software: validations in white literature. http://dx.doi.org/10.2139/ssrn.3783124

[17] Ford, J., (2021). Ethical AI: The Role of Law and Regulation. *Available at SSRN 3831211*. DOI: 10.2139/ssrn.3831211

[18] Franzke, A.S., (2022). An exploratory qualitative analysis of AI ethics guidelines. *Journal of Information, Communication and Ethics in Society*, *20*(4), pp.401-423. DOI: 10.1108/jices-12-2020-0125

[19] Georgieva, I., Lazo, C., Timan, T. and van Veenstra, A.F., (2022). From AI ethics principles to data science practice: a reflection and a gap analysis based on recent frameworks and practical experience. *AI and Ethics*, *2*(4), pp.697-711. DOI: 10.1007/s43681-021-00127-3

[20] Hagendorff, T., (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and machines*, *30*(1), pp.99-120. DOI: 10.1007/s11023-020-09517-8

[21] Hardman, L., (2022). Cultural Influences on Artificial Intelligence: Along the New Silk Road. *Perspectives on Digital Humanism*, pp.233-239. DOI: 10.1007/978-3-030-86144-5_31

[22] Jameel, T., Ali, R. and Toheed, I., (2020). Ethics of artificial intelligence: research challenges and potential solutions. In *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 1-6). IEEE. DOI: 10.1109/iCoMET48670.2020.9073911

[23] Kassens-Noor, E., Siegel, J. and Decaminada, T., (2021). Choosing Ethics Over Morals: A Possible Determinant to Embracing Artificial Intelligence in Future Urban Mobility. *Frontiers in Sustainable Cities*, *3*, p.723475. DOI: 10.3389/frsc.2021.723475

[24] Kazim, E. and Koshiyama, A.S., (2021). A high-level overview of AI ethics. *Patterns*, *2*(9). DOI: 10.1016/j.patter.2021.100314

[25] Khan, A.K., (2022). Trust in Artificial Intelligence: Toward Measuring the Impact of Public Perception. https://aisel.aisnet.org/treos_amcis2022/54

[26] Khan, W. and Fadziso, T., (2020). Ethical Issues on Utilization of AI, Robotics and Automation Technologies. *Asian Journal of Humanity, Art and Literature*, *7*(2), pp.79-90. https://doi.org/10.18034/ajhal.v7i2.521

[27] Kim, M. and Hong, B., (2022). Nursing Ethical Considerations in the AI-Based Technologies. *Robotics & AI Ethics*, *7*(2), pp.10-21. DOI: 10.22471/ai.2022.7.2.10

[28] Leimanis, A., (2020). Self-Imposed Ethical Guidelines For Ai In Education. In *International Scientific Conference" Society, Integration, Education-SIE2020"*.

[29] Liaw, S.T., Liyanage, H., Kuziemsky, C., Terry, A.L., Schreiber, R., Jonnagaddala, J. and de Lusignan, S., (2020). Ethical use of electronic health record data and artificial intelligence: recommendations of the primary care informatics working group of the international medical informatics association. *Yearbook of Medical Informatics*, *29*(01), pp.051-057. DOI: 10.1055/s-0040-1701980

[30] Lubin, A., (2022). Big Data and the Future of Belligerency: Applying the Rights to Privacy and Data Protection to Wartime Artificial Intelligence. *Handbook on Warfare and Artificial Intelligence (Geiss & Lahmann eds., 2022 Forthcoming), Indiana Legal Studies Research Paper*, (460). https://ssrn.com/abstract=3919195

[31] Lysaght, T., Lim, H.Y., Xafis, V. and Ngiam, K.Y., (2019). AI-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research. *Asian Bioethics Review*, *11*, pp.299-314.

[32] Madhavan, R., Kerr, J.A., Corcos, A.R. and Isaacoff, B.P., (2020). Toward trustworthy and responsible artificial intelligence policy development. *IEEE Intelligent Systems*, *35*(5), pp.103-108. DOI: 10.1109/MIS.2020.3019679

[33] Maharani, S.D., Amin, K. and Taufiqoh, A.F., (2022). Technological Progress, Artificial Intelligence Development and Ethical Paradigms. *Wisdom*, (2 (22)), pp.104-117. DOI: 10.24234/wisdom.v22i2.636

[34] Mantini, A., (2022). Technological sustainability and artificial intelligence algor-ethics. *Sustainability*, *14*(6), p.3215. DOI: 10.3390/su14063215

[35] Mezgár, I., (2021). From Ethics to Standards; an Overview of AI Ethics in CPPS. *IFAC-PapersOnLine*, *54*(1), pp.723-728. DOI: 10.1016/j.ifacol.2021.08.084

[36] orstad, K.T., (2020). Intersection of artificial intelligence and medicine: tort liability in the technological age. *Journal of Medical Artificial Intelligence*, *3*.

[37] Panait, C., Ljubenkov, D. and Alic, D., (2021). Striking the balance between innovation and regulation in AI-is Europe leading the way or lagging behind?

[38] Park, G. and Bae, M., (2022). A Case Study on Current Issues in Artificial Intelligence and its' Ethical Implications. *Robotics & AI Ethics*, *7*(2), pp.47-56. DOI: 10.22471/ai.2022.7.2.47

[39] Pasricha, S., (2022). AI Ethics in Smart Healthcare. *IEEE Consumer Electronics Magazine*. DOI: 10.1109/MCE.2022.3220001

[40] Peters, D., Vold, K., Robinson, D. and Calvo, R.A., (2020). Responsible AI—two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, *1*(1), pp.34-47. DOI: 10.1109/TTS.2020.2974991Wyeld, T.G., Leavy, B., Carroll, J., Gibbons, C., Ledwich, B. and Hills, J., (2007). The Ethics of Indigenous Storytelling: using the Torque Game Engine to Support Australian Aboriginal Cultural Heritage. In *DiGRA Conference* (pp. 261-268).

[41] Puyol-Antón, E., Ruijsink, B., Piechnik, S.K., Neubauer, S., Petersen, S.E., Razavi, R. and King, A.P., (2021). Fairness in cardiac MR image analysis: an investigation of bias due to data imbalance in deep learning based segmentation.

In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24* (pp. 413-423). Springer International Publishing. DOI: 10.1007/978-3-030-87199-4_39

[42] Rakova, B., Yang, J., Cramer, H. and Chowdhury, R., (2021). Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), pp.1-23. https://doi.org/10.1145/3449081

[43] Rodrigues, R., (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, *4*, p.100005.Stahl, B.C. and Stahl, B.C., (2021). Ethical issues of AI. *Artificial Intelligence for a better future: An ecosystem perspective on the ethics of AI and emerging digital technologies*, pp.35-53. DOI: 10.1007/978-3-030-69978-9_4

[44] Ryan, M., Antoniou, J., Brooks, L., Jiya, T., Macnish, K. and Stahl, B., (2021). Research and practice of AI ethics: a case study approach juxtaposing academic discourse with organisational reality. *Science and Engineering Ethics*, *27*, pp.1-29. DOI: 10.1007/s11948-021-00293-x

[45] Sanusi, I.T. and Olaleye, S.A., (2022). An insight into cultural competence and ethics in K-12 artificial intelligence education. In *2022 IEEE global engineering education conference (EDUCON)* (pp. 790-794). IEEE. DOI: 10.1109/EDUCON52537.2022.9766818

[46] Siqueira de Cerqueira, J.A., Acco Tives, H. and Dias Canedo, E., (2021). Ethical guidelines and principles in the context of artificial intelligence. In *XVII Brazilian Symposium on Information Systems* (pp. 1-8). DOI: 10.1145/3466933.3466969

[47] Stark, L. and Hoey, J., (2021). The ethics of emotion in artificial intelligence systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 782-793). https://doi.org/10.1145/3442188.3445939

[48] Steinke, G., LaBrie, R. and Sarkar, S., (2022). Recommendation for Continuous Ethical Analysis of AI Algorithms. In *Proceedings of the 2022 European Interdisciplinary Cybersecurity Conference* (pp. 104-106). DOI: 10.1145/3528580.3532996

[49] Subramanian, R., (2017). Emergent AI, social robots and the law: Security, privacy and policy issues. *Subramanian, Ramesh (2017)" Emergent AI, Social Robots and the Law: Security, Privacy and Policy Issues," Journal of International, Technology and Information Management*, *26*(3). DOI: 10.58729/1941-6679.1327

[50] Taddeo, M., McCutcheon, T. and Floridi, L., (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence*, *1*(12), pp.557-560.

[51] Tonidandel, S., King, E.B. and Cortina, J.M., (2018). Big data methods: Leveraging modern data analytic techniques to build organizational science. *Organizational research methods*, *21*(3), pp.525-547. https://doi.org/10.1177/1094428116677299

[52] Umbrello, S., (2022). The role of engineers in harmonising human values for AI systems design. *Journal of Responsible Technology*, *10*, p.100031. DOI: 10.21203/rs.3.rs-709596/v1

[53] Vousinas, G.L., Simitsi, I., Livieri, G., Gkouva, G.C. and Efthymiou, I.P., (2022). Mapping the road of the ethical dilemmas behind Artificial Intelligence. *Journal of Politics and Ethics in New Technologies and AI*, *1*(1), pp.e31238-e31238. https://doi.org/10.12681/jpentai.31238

[54] Wei, M. and Zhou, Z., (2022). Ai ethics issues in real world: Evidence from ai incident database. *arXiv preprint arXiv:2206.07635*. DOI: 10.48550/arXiv.2206.07635Levantino, P.F., (2020). *Framing the picture: A human rights-based study on AI, the case of Facial Recognition Technology* (Doctoral dissertation).

[55] Yigitcanlar, T., Degirmenci, K. and Inkinen, T., (2022). Drivers behind the public perception of artificial intelligence: insights from major Australian cities. *AI & society*, pp.1-21. DOI: 10.1007/s00146-022-01566-0

[56] Zhou, N., Zhang, Z., Nair, V.N., Singhal, H. and Chen, J., (2022). Bias, fairness and accountability with artificial intelligence and machine learning algorithms. *International Statistical Review*, *90*(3), pp.468-480. DOI: 10.1111/insr.12492Zhou, N., Zhang, Z., Nair, V.N., Singhal, H., Chen, J. and Sudjianto, A., (2021). Bias, Fairness, and Accountability with AI and ML Algorithms. *arXiv preprint arXiv:2105.06558* https://doi.org/10.48550/arXiv.2105.06558