World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Ethics of artificial intelligence: Examining moral accountability in autonomous decision-making systems

Jin young Hwang *

University of Edinburgh MA Social Policy and Economics, United Kingdom.

World Journal of Advanced Research and Reviews, 2024, 23(03), 3192-3198

Publication history: Received on 08 August 2024; revised on 19 September 2024; accepted on 22 September 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.23.3.2884

Abstract

This research critically explores the ethical challenges posed by autonomous artificial intelligence (AI) systems, focusing on the moral accountability of decision-making processes conducted without human oversight. Autonomous systems, with applications in healthcare, finance, transportation, and military domains, challenge traditional ethical frameworks such as deontology, utilitarianism, and virtue ethics. By examining these systems' capacity to make decisions with profound societal impacts, the study addresses the growing tension between algorithmic decision-making and established notions of human moral responsibility. Key topics include the "moral machine problem," where AI systems face ethical dilemmas in life-or-death scenarios, and the role of algorithmic bias, which can perpetuate inequality and harm. The research evaluates existing accountability mechanisms, highlighting their limitations in addressing the ethical and legal complexities introduced by AI. Furthermore, it examines alternative frameworks, such as relational ethics and collective responsibility, which emphasize shared accountability among developers, users, and societal stakeholders. The study proposes practical strategies for embedding ethical principles into AI design, advocating for increased transparency, explainability, and oversight. It argues that while traditional philosophical theories provide valuable insights, they must be adapted to address the unique challenges of AI systems. By integrating these insights with contemporary technological realities, this research contributes to the ongoing discourse on ensuring ethical and accountable AI deployment, ultimately seeking to align technological advancement with societal values and human welfare.

Keywords: Artificial intelligence; Moral accountability; Algorithmic bias; Deontology; Utilitarianism; Autonomous systems

1. Introduction

1.1. Background and Context

Artificial Intelligence is the design of computer systems that can perform human tasks such as visual perception, decision making/problem solving, and understanding natural language. AI systems can be divided into two main types: Narrow or weak AI, designed to perform a particular task (e.g., facial recognition or voice assistants), and general or strong AI, which is human-like in its abilities and can solve many problems on its own (Russell and Norvig, 2020). AI is advancing very rapidly nowadays, especially in autonomous systems where machines make decisions without a human being present.AI-aided automated decision-making systems have applications in several fields. AI-powered healthcare systems diagnose diseases or suggest treatment based on large data sets (Russell and Norvig, 2020). In finance, trends in markets are predicted using algorithms, trading is automated, and risk is assessed. The transportation sector has seen autonomous vehicles with AI making driving decisions in real time. In the military, autonomous weapons and drones now make life-or-death decisions without humans being involved. Such advances create enormous

* Corresponding author: Jin young Hwang

opportunities but also pose huge ethical questions regarding moral responsibility when AI systems make decisions that affect millions of people worldwide. The liability issue arises especially with autonomous AI systems. Human agents traditionally make decisions, but AI systems break this framework by acting autonomously without human oversight. And so the question is: Who is liable when autonomous AI makes bad or good decisions? Is it the developer who programmed the system, the user who deployed it, or is it the AI system itself? This questions fundamental philosophical issues such as responsible action and morality: How can well-established ethical theories such as deontology, utilitarianism, and virtue ethics be adapted or applied to AI decision-making? Viewing AI through the prism of traditional philosophical theories helps to understand how responsibility is assigned. As ethics is concerned with obligations and rules, some found it hard to accept that machines make decisions using programmed algorithms. If the greatest good is their goal, then evaluating AI systems without human reason or emotion may prove problematic. As machines have no moral character like humans, ethics – focusing on who makes decisions based on their own traits – is difficult to apply to AI (Russell and Norvig, 2020). These philosophical inquiries have become essential for responsible development, deployment, and management of AI systems.

1.2. Research Purpose and Objectives

This study aims at critically analyzing ethical implications of AI decision-making systems and how these systems challenge classical philosophical theories on moral responsibility. With AI developing and becoming more embedded in everyday life, responsibility has to be understood and assigned to it. This study examined how the autonomous character of AI calls for a reconfiguration of accountability frameworks.

The objectives of the study are

- Explore moral responsibility in AI systems: This includes analyzing how autonomous AI systems challenge moral responsibility attribution.
- Examine whether traditional philosophical theories apply to AI decision-making or not: This work examines how deontology, utilitarianism, and virtue ethics address special features of AI systems in complex real-world situations where human decision-making is replaced by algorithms.
- Discuss how to create new accountability frameworks: Given the limitations of existing ethical theories, the study explored how to adapt these frameworks to the particular challenges of autonomous AI systems.

1.3. Research Questions

These research questions guide this study

- Primary Question: What applications can traditional philosophical theories of responsibility such as deontology and utilitarianism find for the moral responsibility of autonomous AI systems?
- Secondary Question: How do AI systems challenge establish moral responsibility concepts? How do we assign blame or praise when AI systems make decisions in complex or life-or-death situations? Which ethics frameworks should account for the particular features of autonomous AI systems?

2. Literature Review

2.1. Understanding Autonomous AI Systems

Autonomous AI systems make decisions and perform tasks without human intervention. They use complicated algorithms, machine learning, and huge data sets to perform tasks that humans usually do. Automotive systems have many applications in transportation, healthcare, finance, and even military domains (Binns, 2020). Autonomous cars, medical diagnostic tools, financial trading algorithms, and autonomous drones are examples of such systems. Such systems work by sensing their environment, processing information, and acting according to learned or pre-determined patterns often optimized to meet predetermined goals (Binns, 2020).

Types of Autonomous AI Systems

- Autonomous automobiles: Driven by AI, these vehicles can drive in traffic, make real-time decisions, and respond to unexpected events or road conditions. These systems drive safely with deep learning, sensor data like LIDAR, cameras, and complex algorithms (Carlo, 2019).

- AI in healthcare: In medicine, AI systems process patient data to recommend treatments or preventative care. These AI models learn from large data sets, but ethical questions arise regarding trust, privacy, and accountability (Bryson and Winfield, 2020)Calo.
- Military applications: Drones and autonomous weapon systems can independently target and carry out missions, posing questions about war ethics and risks of accidental harm.
- Financial algorithms: AI is applied to trading and investing, taking decisions based on mathematical models and huge amounts of data to forecast market trends and optimize portfolios.

Despite their growing capabilities, autonomous AI systems have limitations such as cognitive issues, decision bias, and reliance on incomplete or biased data (Carlo, 2019). These limitations often create ethical quandaries, as in the case of autonomous cars in collision scenarios where the AI must choose who or what to damage to minimize harm. This "moral machine problem" demonstrates one of the moral problems of AI: Can an algorithm be programmed to make ethical decisions, or should humans make such decisions?

3. Methodology

3.1. Research Design

This study adopts a qualitative research methodology, utilizing a combination of theoretical analysis and case studies. The qualitative approach is appropriate for exploring the complex ethical implications of AI decision-making systems, as it enables an in-depth examination of philosophical theories and real-world applications. By analyzing existing literature and evaluating case studies, this research aims to identify gaps in current ethical frameworks and propose recommendations for addressing accountability challenges posed by autonomous AI systems.

3.2. Data Collection

Data for this study was collected from secondary sources, including academic journals, books, conference proceedings, and credible online publications. The selection criteria focussed on sources that address AI ethics, moral accountability, and related philosophical theories. Key case studies, such as incidents involving autonomous vehicles and military drones, was also analyzed to provide practical insights into the ethical dilemmas associated with AI decision-making.

3.3. Data Analysis

The collected data was analyzed through thematic analysis, which involves identifying, organizing, and interpreting patterns or themes within the data. This method is suitable for examining the intersection of philosophical theories and practical applications of AI ethics (Boddington, 2020). Themes such as "moral agency," "algorithmic bias," and "distributed responsibility" guided the analysis, enabling a structured exploration of how traditional ethical frameworks can be adapted to AI systems.

3.4. Ethical Considerations

Given the focus on ethical accountability, this research adheres to high ethical standards by ensuring that all sources are properly cited and that the analysis is conducted with academic integrity. Furthermore, the study does not involve human participants, thus eliminating concerns related to consent and privacy. The ethical implications of the proposed recommendations were also be critically evaluated to ensure their feasibility and alignment with societal values.

3.5. Limitations

This study is limited by its reliance on secondary data, which may not fully capture the nuances of emerging AI technologies. Additionally, the analysis is constrained by the availability of case studies and the inherent challenges of applying philosophical theories to rapidly evolving technological contexts. Despite these limitations, the research provides valuable insights into the ethical challenges of AI systems and contributes to the broader discourse on moral accountability in autonomous decision-making.

4. Data Analysis, Presentation and Interpretation

4.1. Revising Existing Philosophical Models for AI Accountability

AI technology has changed traditional philosophical models of accountability. Philosophical foundations like deontology, utilitarianism, and virtue ethics offer some insight but do not address the difficulties of autonomous

systems. AI making decisions autonomously often in dynamic settings where moral dilemmas arise challenges established notions of responsibility (Greene et al., 2019). These frameworks should therefore be modified or enlarged to take into account AI systems' specific features.

A key change is recognizing that responsibility for AI is not just about assigning blame to some agent but the whole system of actors creating, deploying, and using AI systems. The traditional models of accountability were constructed when human actors alone made the decisions. With autonomous AI systems, responsibility must be distributed rather than singular (Gabriel, 2020). This is particularly true in "shared responsibility" models—in which the activities of an AI are given to developers, users, policymakers, and regulators.

With the distributed responsibility model, AI developers are liable for ethical design of the system—free of bias and harmful effects (Obermeyer and Emanuel, 2019). They should code ethical principles into the AI. Users of the technology also share responsibility, since their actions may affect how an AI system is deployed/used in practice. With autonomous vehicles, for example, the human driver might still be expected to watch the system and intervene when required. Regulators ensure the AI is legal and ethical in all areas—particularly in high-risk areas such as military applications and healthcare (Obermeyer and Emanuel, 2019).

This is a departure from human-centered responsibility models of the past that recognize accountability must have multiple layers of oversight. Because AI behaviors in real-world applications are unpredictable—for example, self-driving cars may fail to avoid an accident or biased decision-making in hiring algorithms may be unethical—a framework that accepts collective responsibility rather than individual blame seems more appropriate (Jobin et al., 2019). Here shared accountability is required since no one can predict or control how autonomous AI systems act (Obermeyer and Emanuel, 2019).

Modifying traditional ethical frameworks means looking at AI systems as tools or moral agents that make autonomous decisions instead of as dynamic components of a system of responsibility. AI systems are now cooperative systems within sociotechnical systems that require collective oversight and accountability.

4.2. The Role of AI in Human Society and the Need for Ethical Oversight

AI technology has increasingly ethical implications as it is incorporated in society. Artificial Intelligence systems are being applied in healthcare, military, transport, and public services—sectors that could potentially affect human lives. The increasing reliance on AI in such crucial areas demands new ethical frameworks so that AI systems consider human welfare, autonomy, and well-being as their primary concern (Eubanks, 2020).

In healthcare, AI applications can be diagnostic tools or robotic surgeries posing questions regarding patient privacy, informed consent, and error possibility. In military applications, AI has been applied to autonomous weapons systems and decision-making, bringing up doubts about the morality of handing life-or-death decisions to computers. And in transportation, AI in self-driving cars raises questions of safety, accountability, and how AI systems should act in emergency situations (Eubanks, 2020).

The implications for society require human-centric ethical oversight of AI. AI systems should aim at human flourishing, respect autonomy, and minimize harm. Developers, policymakers, and ethicists have to work in concert to make certain AI systems don't violate human rights or dignity but advance human abilities and serve the common good (Dignum, 2019).

A global cooperation on AI ethics is also needed. AI technologies are always transnational, so ethics guidelines and standards must be developed worldwide. Countries must agree on norms so AI can develop for humanity but without introducing discrimination, surveillance, and privacy risks (Moor, 2021.

4.3. Policy and Legal Implications for AI Accountability

Rapid deployment of AI technologies outshone appropriate policies and regulations. Comprehensive legal frameworks are needed to design and use AI systems in a human rights and societally beneficial manner. These frameworks should address issues including transparency, bias, privacy, and accountability and hold individuals and organizations responsible when AI systems cause harm (Wachter et al., 2019).

Current legal approaches to AI accountability are evolving and have gaps. For example, tort law and product liability are not prepared for AI challenges when autonomous systems act in unpredictable or harmful ways. Regulated activities often do not take into account the complexity of AI decision-making where the line between human and machine action

blurs. A third issue is liability—who pays when an AI makes a bad decision? Who owns the AI—the developer, the user, or the company?

To safeguard human rights and societal welfare, legal systems need to clarify AI accountability frameworks. This includes changing existing laws to accommodate autonomous systems—like the "black box problem" where AI makes its decisions opaque. In addition, new policies must ensure that AI systems are transparent and explicable—that is, people and regulators understand how decisions are made (Ashrafian, 2019).

Hence, AI can improve many aspects of human life, but its increasing incorporation into society calls for a new conception of accountability and ethical responsibility. Legal frameworks have to be adapted to AI challenges so that technologies developed and used benefit humanity without causing harm. Robust laws and regulations were needed to keep AI systems ethical and accountable to society.

4.4. Analysis of Ethical, Philosophical, and Policy Implications

4.4.1. Integrating Ethical Theories with Technological Realities

AI in complex sociotechnical systems poses new dimensions of accountability and moral responsibility that challenge established ethical frameworks. Classical theories like deontology, utilitarianism, and virtue ethics have to be rethought in light of AI features such as decision ambiguity ("black box problem") and lack of intrinsic moral intent. This disconnect requires an evolution of these frameworks rather than their outright dismissal. For example, deontology's rule-based principles may be extended to include compliance monitoring systems that ensure AI meets pre-defined ethical constraints, and utilitarianism may guide the assessment of AI's outcomes through stakeholder-oriented evaluation models (Wachter et al., 2019).

Yet the biggest obstacle lies in connecting theoretical ethics to practice. Many AI applications operate in areas where ethics often conflict—for example, patient confidentiality versus benefits of sharing data in healthcare (Wachter et al., 2019). As such, the analysis suggests that hybrid ethical models are needed which combine the rigidity of traditional frameworks with the flexibility needed to deal with real-world problems. Such an approach stresses co-responsibility—developers, users, and regulators jointly manage AI's ethical dimensions.

4.4.2. Distributed Accountability: A Paradigm Shift

Analyzing distributed responsibility reveals a paradigm shift towards collective accountability. Although traditional notions of accountability are anthropocentric, the autonomous nature of AI requires a broader lens that takes into account AI ecosystem interdependencies (Shneiderman, 2020). For example, in autonomous vehicles, the accountability extends beyond manufacturers and developers to end-users who may share liability with respect to their interaction with the system.

With this distributed model, accountability becomes a continuum instead of a binary. Developers should build safeguards and ethical considerations into the design phase, while regulators check that the systems meet societal norms (Shneiderman, 2020). Also, end users need to understand the limits of AI autonomy and their role in monitoring its operations. The analysis suggests distributed accountability is a response to AI's complexity as well as a proactive framework for reducing ethical risks.

4.4.3. Ethical Implications of Global AI Governance

The analysis calls for global cooperation on AI ethics and governance as AI technologies are intrinsically transnational. Unstandardized ethical guidelines could create regulatory arbitrage where developers and companies exploit jurisdictions with lax ethical oversight. Lacking a unified ethical framework risks increasing global inequalities as AI enters new frontiers of predictive policing, surveillance, and labor automation (Shneiderman, 2020).

Ethical standards need collaborative efforts such as international agreements or partnerships. They should address universally accepted principles such as non-discrimination, privacy, and transparency while recognizing cultural and contextual differences. For example, ethical questions regarding bias in AI decision-making could be addressed globally by requiring transparency of training data and algorithms (Shneiderman, 2020). Such a foundation would allow fair AI development and deployment and build trust across borders.

4.4.4. Legal Implications and Challenges

The analysis shows grave deficiencies in the present legal systems, particularly on liability and transparency fronts. While distributed accountability models seek to accommodate the multi-stakeholder nature of AI, legal systems must adapt to this complexity. For instance, liability is not resolved when AI systems operate autonomously and produce undesirable outcomes. Typical legal frameworks that focus on human actions struggle to assign responsibility when harm results from AI decisions (Shneiderman, 2020).

This gap highlights the need for legal innovation such as the creation of controversial "AI personhood" concepts for legal and liability purposes. Or a pragmatic solution might be extending existing regulatory frameworks to include mandatory insurance for AI systems. For example, manufacturers of autonomous vehicles might be required to carry liability insurance that covers victims irrespective of fault—making the legal process simpler and ensuring accountability (Shneiderman, 2020).

4.4.5. Policy Recommendations: Dynamic and Adaptive Governance

Effective governance needs adaptable policies that look forward. Rapidly evolving AI requires constant updates to regulations. For instance, policy frameworks should provide for algorithm audits and require AI developers to disclose and justify decision-making. These would increase accountability while addressing AI opacity (Lepri et al., 2018).

The analysis also calls for inclusive policymaking. Public engagement—for example, through citizen panels or consultations—can reveal much about society's expectations and concerns regarding AI technologies. Policies also need to ensure that underrepresented groups—often disproportionately impacted by AI biases—are represented in setting ethical and regulatory standards (Lepri et al., 2018).

4.4.6. Synthesis and Reflection

Effective governance needs adaptable policies that look forward. Rapidly evolving AI requires constant updates to regulations. For instance, policy frameworks should provide for algorithm audits and require AI developers to disclose and justify decision-making. These would increase accountability while addressing AI opacity.

The analysis also calls for inclusive policymaking. Public engagement—for example, through citizen panels or consultations—can reveal much about society's expectations and concerns regarding AI technologies (Mulligan and Bamberger, 2019). Policies also need to ensure that underrepresented groups—often disproportionately impacted by AI biases—are represented in setting ethical and regulatory standards.

5. Conclusion

The advent of autonomous AI systems has raised philosophical questions that traditional ethical theories fail to address. AI as a moral agency, a moral accountability, and a moral responsibility raise fundamental questions about moral agency/accountability/responsibility that AI poses as it can make decisions autonomously without human intervention. Traditional models like deontology, utilitarianism, and virtue ethics provide some insight but fail to account for AI's particular features (Danaher, 2020). Lacking moral intention in AI systems prevents deontological ethics from being applied, utilitarian AI decision-making in life-or-death scenarios raises doubts about the limit of outcome-based ethical reasoning. Also, virtue ethics based on human character does not translate well into AI systems that do not actually embody virtues. AI's autonomy and potential for autonomous decision-making demand new ethical frameworks that integrate human responsibility at multiple levels. These new frameworks like distributed responsibility call for shared accountability among AI developers/users/regulators/society. These frameworks recognize AI's unpredictability and human welfare implications and call for collaborative oversight and multidisciplinary approaches to AI governance. These findings highlight the ethical challenges involved in the design and deployment of autonomy AI systems. These systems' ability to make decisions autonomous of human input raises deep philosophical questions about moral agency, accountability, and responsibility. Conventional ethical frameworks like deontology, utilitarianism, and virtue ethics offer fundamental insights but are inadequate in addressing the AI-specific capabilities and limitations. Deontological ethics struggles with AI's lack of intentionality whereas utilitarian ethics is confronted with ethical dilemmas in life-or-death decision-making and reduces morality to a simplified calculus. As well, virtue ethics is always anthropocentric and based on human qualities that AI lacks. As AI systems become autonomous, ethical reasoning must take a new direction towards distributed responsibility. This framework places an emphasis on a shared ethical accountability for AI developers/users/regulators/society given the collective nature of AI creation/deployment/impact. With this distributed model, ethical oversight is ensured throughout the lifecycle of AI systems, from multiple perspectives to limit risks and maximize social benefits.

Compliance with ethical standards

Statement of ethical approval

Ethical approval was obtained.

References

[1] Ashrafian, H. (2019). AI as the Next Global Public Health Challenge: Are We Ready?. Health Policy and Technology, 8(2), 101-103.

[2] Binns, R. (2020). On the Apparent Conflict Between AI Ethics and Accountability. Philosophical Transactions of the Royal Society A, 378(2164), 20190286.

[3] Boddington, P. (2020). Towards a Code of Ethics for Artificial Intelligence. Springer.

[4] Bryson, J. J., & Winfield, A. F. T. (2020). Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems. Computer, 53(10), 18-23.

[5] Calo, R. (2019). Artificial Intelligence Policy: A Primer and Roadmap. UC Davis Law Review, 51(2), 399-435.

[6] Danaher, J. (2020). Automation and Utopia: Human Flourishing in a World Without Work. Harvard University Press.

[7] Dignum, V. (2019). Responsible Artificial Intelligence: Designing AI for Human Values. Artificial Intelligence, 280, 103207.

[8] Eubanks, V. (2020). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

[9] Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review, 1(1).

[10] Gabriel, I. (2020). Artificial Intelligence, Values, and Alignment. Minds and Machines, 30(3), 411-437.

[11] Greene, D., Hoffmann, A. L., & Stark, L. (2019). Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning. Proceedings of the 52nd Hawaii International Conference on System Sciences, 2122-2131.

[12] Jobin, A., Ienca, M., & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines. Nature Machine Intelligence, 1(9), 389-399.

[13] Lepri, B., Oliver, N., Letouzé, E., Pentland, A. S., & Vinck, P. (2018). Fair, Transparent, and Accountable Algorithmic Decision-Making Processes. Philosophy & Technology, 31(4), 611-627.

[14] Mittelstadt, B. D. (2019). Principles Alone Cannot Guarantee Ethical AI. Nature Machine Intelligence, 1(11), 501-507.

[15] Moor, J. H. (2021). Can AI Systems Be Moral Agents? A Framework for AI Accountability. Ethics and Information Technology, 23(2), 151-165.

[16] Mulligan, D. K., & Bamberger, K. A. (2019). Accountability in Algorithmic Decision-Making: Fairness, Justice, and the Law. Annual Review of Law and Social Science, 15, 1-17.

[17] Obermeyer, Z., & Emanuel, E. J. (2019). Predicting the Future—Big Data, Machine Learning, and Clinical Medicine. New England Journal of Medicine, 381(12), 1145-1154.

[18] Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson Education.

[19] Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Reliable, Safe, and Trustworthy. International Journal of Human-Computer Interaction, 36(6), 495-504.

[20] Wachter, S., Mittelstadt, B., & Floridi, L. (2019). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76-99.