International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(REVIEW ARTICLE)

Check for updates

The ethics of AI decision-making: Balancing innovation and accountability

Apoorva Kasoju * and Tejavardhana Vishwakarma

1228 170TH ST SW, Unit A, Lynnwood, WA, 98037, USA.

International Journal of Science and Research Archive, 2024, 12(02), 3084–3095

Publication history: Received on 11 July 2024; revised on 17 August 2024; accepted on 20 August 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.12.2.1548

Abstract

Attention from public domains and academic circles has intensified due to the way AI systems' decision-making enters various societal structures of healthcare, criminal justice, finance, and autonomous technologies. This paper explores how technology innovation and accountability aspects interlock in AI decision-making systems through examination of ethical structures together with regulatory voids and problems resulting from AI system implementations. The paper examines the decision-making processes of present AI systems while exploring auditability together with Explainability measures and determining liability when mistakes or discriminatory operations occur.

The paper employs three case examples from facial identification systems, predictive crime analysis, and medical technology to prove how unrestrained development methods could sustain discrimination patterns while breaking public confidence. We present an ethical AI governance framework which depends on making the system visible and requiring human intervention together with consideration of specific circumstances. The study finds that ethical development of AI systems demands more than performance and efficiency since it requires the implementation of legally enforceable social accountability mechanisms. Social and economic systems require immediate mutual effort to develop AI systems that support human values and democratic principles.

Keywords: Artificial Intelligence (Ai); Ethical Decision Making; Algorithmic Accountability; Ai Governance; Transparency.

1. Introduction

Modern society has transformed the concept of Artificial Intelligence, which was once considered futuristic, into a daily practice that penetrates healthcare, criminal justice, transportation, finance systems, and educational institutions. AI systems currently affect decisions that humans traditionally handled, by operating both as conversational simulation chatbots and parole outcome-determining algorithms. These high-efficiency technologies present complicated social, as well as moral dilemmas that emerge when they play a role in critical choices. Advanced machine learning and deep learning algorithms are rapidly enhancing decision-making transformation, which requires deep examination of machine involvement in human life decisions and their decision-making parameters.

Issues such as AI bias, transparency, and accountability have become major public concerns for public discussion during the past few years. The public found cause for concern when software flaws in facial identification and predictive policing tools showed racial bias, credit scoring algorithms operated with unclear decision-making processes, leading to problems of unfairness, discrimination, and systemic inequality. AI development must be studied both through technical system operation analysis and through evaluations of ethical frameworks and deployment management systems.

* Corresponding author: Apoorva Kasoju

1.1. Problem Statement

The tremendous value of AI faces challenges as high-risk decisions now utilize AI systems before adequate ethical rules and protective measures can be established. Decision-making systems that implement AI technology enter public service before dedicated assessments evaluate their societal effects or identify potential future consequences from their operational decisions. The growing adoption of black box models presents the most concerning issue, since users usually cannot understand how systems arrive at their decisions. The absence of definite responsibility emerges when such systems fail or create biased results.

Current legal and ethical systems lack the ability to address AI-made decisions with proper resolution. Traditional human-based accountability methods fail to function with AI systems because artificial intelligence presents unclear connections between decision-makers and operates through autonomous or probabilistic systems. The inability to monitor and understand AI decision-making processes leads people to develop skepticism towards these technologies, while impeding their appropriate usage.

1.2. Objectives and Research Questions

The paper investigates ethical problems with machine decision systems while examining the balance between innovation and system accountability in these algorithms. The paper presents an interdisciplinary approach to handle these matters by integrating expertise from computer science, ethical principles, legal frameworks, and public policy frameworks. The primary inquiry of this research investigates two main questions.

- Current AI systems use what methods to make critical decisions while their ethical responsibilities to such decisions remain uncertain.
- What forms do biases appear in AI systems and which detection methods exist to eliminate these biases?
- Multiple models of governance operate in the present or new ones need creation for the responsible use of AI systems.
- The implementation of transparency and Explainability features in AI systems should occur without damaging operational performance.
- This paper supports current discussions regarding AI governance principles for ethical innovation by addressing important questions about future AI management and ethical research guidelines.

1.3. Scope and Significance

The study concentrates on examining AI implementations used for critical decision operations that strongly affect both community members and individual citizens. Facial identification tools used by law enforcement personnel, predictive technology within criminal sentencing, and diagnostic computational systems in medical settings comprise the studies examined. The selected examples have been chosen because they matter at the societal level and reveal fundamental ethical and legal implications.

This study carries substantial importance because it aims to direct Artificial Intelligence development and management frameworks. This paper bases its forward-thinking remarks on ethical algorithmic solutions by emphasizing transparency, human oversight, and accountability models. The research actively works to unify technological progress with moral analysis, as AI increasingly defines our social structure and political operations.

1.4. Structure of the Paper

1.4.1. The paper adopts the following organization

In this section, the paper performs a review of academic and industry literature about AI ethics, algorithmic bias, accountability frameworks, and governance structures. The methodology section provides details on the research approach, which integrates interdisciplinary methods for case study selection, qualitative evaluation processes, and ethical evaluation techniques. This segment demonstrates findings and identifies main issues and consistent patterns in artificial intelligence decision-making through selected case examples. The fifth section analyzes discoveries through examination of present ethical frameworks and investigates future adjustments in the field. The paper's main contributions and research and policy development recommendations

2. Literature review

2.1. Ethical foundation of AI

The development of artificial intelligence follows traditional philosophical ethics but brings unknown difficulties when applied to computational systems. AI systems are assessed using utilitarianism and deontology as traditional moral frameworks to determine their conduct. The process of integrating these conceptual ethical principles into AI technical systems stands as an overwhelming technological difficulty. Nick Bostrom and ethicists Wendell Wallach, along with Colin Allen, investigate how machines can achieve moral decision-making ability through their research on machine morality. The foundation of this concept depends on AI objectives becoming consistent with human values. Earning a perfect value alignment between machines and humans becomes challenging due to the multiple human ethical perspectives that vary across different cultures. Luciano Floridi, together with Josh Cowls, developed a vast framework for digital ethics which applies biomedical ethical concepts to establish beneficial systems while upholding autonomy principles, following justice guidelines, and preventing harm during AI creation. In practice, the worthwhile principles exist only as theoretical objectives because technical implementation proves difficult to align with ethical theories.

The collected materials suggest that ethical standards battle with actual operational obstacles in practical situations. Higher levels of system complexity and autonomy make it progressively complicated to ensure that AI systems conform to moral standards. The commitment to ethical programming extends beyond coding since it needs organizational principles, stakeholder relations, and social and political elements for delivering responsible AI solutions.

Figure 2 9 Ethical AI Principles For Organization To Follow

2.2. Algorithmic Bias and Fairness

Scientists widely discuss algorithmic bias as the technical problem that causes AI systems to copy and intensify current social inequalities. The foundation of this problem exists in the training data used for these systems. The algorithms that derive knowledge from biased historical data will reproduce discriminatory results by default. Barocas and Selbst, along with other scholars, prove that seemingly objective data-based systems maintain institutional prejudice through their operations in credit scoring, law enforcement, and hiring processes. Through her book 'Automating Inequality,' Virginia Eubanks presents ethnographic evidence about AI systems that disproportionately harm minority populations when used in welfare and criminal justice systems. Such technological systems can reinforce patterns of exclusion and invisibility because they usually lack participation from affected communities throughout their development process. Many organizations maintain efficiency and cost-effective approaches, even though they are aware that these priorities overshadow fairness and equity.

Multiple interventions in technical areas have been developed to address algorithmic bias through combinations of fair-based machine learning methods and pre-processing techniques for training data modification. The existing algorithms face limitations because they use poor ethical standards of fairness while maintaining good technical performance. Numerous academics state that statistical decision-making systems are not successful at confronting fairness problems because these should be interpreted by considering broader social and political contexts. The technical solutions offer

beneficial resources, though they lack the necessary power to fully eliminate basic issues of discrimination and injustice across the world.

2.3. Transparency and Explainability

Development of trusted AI systems requires designers to explain their design choices and outline system operational procedures, specifically regarding critical scenarios that affect community members.

In 2020, this field started its journey with transparency as one of its central ongoing difficulties. When deep neural networks (DNNs) implement more complex algorithms and machine learning models, their basic decision systems become progressively difficult to understand. The unclear internal workings of such systems earned the term "black box" from researchers, since users can monitor both inputs and outputs yet remain unaware of internal operations. Researchers Kate Crawford and Jenna Burrell examine algorithmic opacity in their respective work. Burrell demonstrates three distinct explanations for algorithmic opacity: intentional secrecy for protecting proprietary data, user-based illiteracy from technology limitations, and intrinsic opacity due to complex algorithmic models. System opacity creates fundamental challenges for accountability because of its different forms when incorrect or biased decisions occur during system incidents.

XAI has established itself as an active research field that serves to resolve these issues. Users can access simplified decision explanations through local model behavior approximations that LIME and SHAP supply. These methods face criticism because they frequently reduce complex model operations into smaller simplified forms which deviate from actual model operation. The models remain obscure for non-technical evaluators, which reduces their utility during public or legal exposure sessions. Explainable AI requires proper consideration of audience-specific messaging approaches, model development openness, and ethical evaluation of interpretive processes.

Table 1 Comparative Overview of Key Approaches to AI Ethics in Literature and Policy

Source Type	Representative Actor(s)	Framing of Ethics	Focus Areas	Limitations Identified
Academic – Deontological	Floridi, Binns	Ethics as adherence to moral principles and duties	Autonomy, responsibility, transparency	Often abstract; limited operational guidance
Academic – Utilitarian	Russell, Norvig	Ethics as outcome-based decision-making	Harm reduction, safety, alignment	Focuses on idealized scenarios; less attention to inequality
Policy-Oriented – Multistakeholder	OECD, UNESCO, EU Commission	Ethics as balanced, consensus-based governance	Human rights, fairness, accountability	Lacks enforcement; overly generalized principles
Corporate – Self Regulatory	Google, Microsoft, IBM	Ethics as innovation compatible risk management	Bias mitigation, privacy, trust	Ethics used for branding; minimal external oversight
Critical Scholars	Eubanks, Noble, Benjamin	Ethics as Power aware, justice driven critique	Structural injustice, data colonialism, marginalization	Often overlooked in mainstream discourse; calls for systemic change Structural injustice, data colonialism, marginalization

2.4. Accountability and Governance

AI accountability comprises two necessary aspects: responsibility assignment during system failures, and traceable decision-making procedures with auditing capabilities and options for challenge. AI systems constitute exceptional challenges because no single actor handles the decision-making functions, which are spread across multiple actors, including software developers, data providers, vendors, and institutions that operate the technology. When responsibility divides among many actors during interactions between data, algorithms, and institutional practices, it becomes challenging to assign responsibility to a specific person for incurred harm. AI-specific accountability tools need development according to thoughts shared by both legal experts and technological experts. Algorithmic impact assessments represent an integrity measure that mirrors environmental assessments by making organizations responsible for evaluating AI system ethical impacts during pre-deployment evaluations. The General Data Protection Regulation (GDPR) of Europe provides regulations about automated decision-making with a mandate for "right to explanation" protection that acts as legal grounds for algorithmic accountability.

Modern regulatory approaches to technology show delay between innovation cycles and function independently from one another without current standards in place. The current laws suffer limitations because they have unspecified definitions, inadequate enforcement ability, and involve insufficient stakeholder participation. A proactive governance framework needs development because it must be inclusive and adaptable to the changing times. The regulatory frameworks must combine harm regulation with best practices, including participatory system design, independent oversight processes, and institutional openness. The available scholarly research advocates a fundamental change in mindset concerning governance because it needs to serve as a basis for creating sustainable and ethical technological progress.

2.5. Gaps and Emerging Directions

Many essential questions about AI ethics remain unanswered in current research, even though various authors have contributed substantial knowledge to this field. Research about ethical issues in technology focuses primarily on Western regions because investigators use Western legal systems, ethical priorities, and cultural models. The unethical distribution of the literature across specific regions creates a problem, as it dismisses valuable knowledge from different philosophical systems and governance mechanisms. The contemporary literature maintains segregated divisions between technical, philosophical, and legal content because these fields exist independently instead of interacting. The resolution of difficulties surrounding AI decision-making needs genuine interdisciplinary teamwork between computer scientists, ethicists, lawyers, sociologists, and community members who jointly develop solutions. The main concern arises from the fact that applying ethical guidelines functions poorly in real-life situations. Many AI ethical frameworks and guidelines, which include the IEEE's Ethically Aligned Design and the OECD's AI Principles, exist today but they typically remain vague and insufficient for direct implementation. The transformation of ethical principles into actionable strategies depends on developing new assessment systems coupled with institutional procedures which convert values into application methods.

Active research exists for understanding the extensive social effects of AI. Brief studies with limited scope make up most of existing research on prolonged social impacts of AI because they focus on immediate results instead of long-term institutional changes. Research needs to monitor AI evolution because it should examine both institutional and labor patterns evolution as well as public and democratic systems changes. The study fills theoretical gaps by using a combined method of conceptual-based ethical AI decision-making that relies on empirical analysis of case studies.

3. Methodology

3.1. Philosophical Foundations and Epistemological Commitments

The study uses constructivist epistemology combined with critical theory as philosophical foundations. Constructivism acknowledges that new knowledge emerges from how society, cultural factors, and institutional structures construct it. The analysis shows that ethical notions, including accountability and fairness, require context-based environments to understand their meanings when working within artificial intelligence (AI). All ethical decision processes are influenced by the combination of existing power dynamics, organizational standards, and governmental policy decisions. Critical theory from Frankfurt School origins, together with modern studies of science and technology, provide research instruments for understanding the social and political factors within AI systems. The approach refuses to accept technological neutrality because it demands assessments about power distribution, identification of marginalized voices, and the insertion of ideological preferences into systems. Through a conceptual fusion of these different epistemological approaches, the methodology enables a comprehensive evaluation of how AI ethics systems are built and how they become controlled by certain interests.

3.2. Research Design

The study utilizes a qualitative interpretive case study approach to fulfill its purposes. The chosen design method supports the exploration of complex context-dependent situations because it delivers deep understanding of how ethics is implemented within AI systems. The researcher views case studies not as statistical representation but as detailed examples that create knowledge about broader subjects. This research studies real-world AI applications in criminal justice, healthcare, welfare, and surveillance to understand actual institutional, legal, and social dynamics of ethical AI practice, rather than theoretical frameworks. The decision to conduct multiple case studies allows researchers to identify various ways ethical challenges receive interpretation and resolution across different fields. Different governance structures, technological rule sets, and socio-political environmental pressures provide the study framework to assess ethical principle implementation. The case study method does not pursue statistical generalization yet enables researchers to discover conceptual generalizations for future theoretical development and experimental research.

3.3. Data Collection and Source Material

The main method of data collection for this research consists of document-based analysis. This research method uses public discourse and institutional documentation as its foundation because of their abundance regarding AI ethics. This research analyzes multiple textual sources consisting of academic publications, policy reports, ethical guidelines, legal decisions, corporate statements, media investigations, and audit reports. This collection of documents functions as both research evidence and conceptual grounds, which show different stakeholders' attitudes toward ethical matters, their backing of technological actions, and their responses to criticism.

The research analyzes institutional documents as constructs that emerged through conflicting negotiation processes, rather than seeing them as direct behavioral windows. The research team selects documents based on their relevance, their perception as credible sources, and their capacity to present multiple viewpoints, especially by including minority perspectives and critical arguments. Through this method, researchers can recreate ethical narratives about AI while learning about how organizations use ethical values, how these values compete against one another, and how they are both intentionally deployed.

Table 2 Methodological Framework and Research Stages

Stage	Description	Primary Activities	Outcomes
Conceptualization	Framing the research within a constructivist and critical theory paradigm	Identifying philosophical foundations, research questions, and ethical lens	Formation of an interdisciplinary ethical inquiry framework
Case Selection	Choosing real-world examples that reveal ethical tensions in AI systems	Reviewing public controversies, regulatory reports, and institutional use cases	Four case studies reflecting diverse sectors and governance challenges
Corpus Construction	Gathering relevant textual data across institutions and discourses	Collecting policy documents, academic texts, corporate statements, audits, and media	Comprehensive textual database for qualitative analysis
Thematic Coding	Extracting key themes and ethical patterns across the data set	Iterative close reading and open coding using interpretive methods	Identification of dominant ethical concepts and contradictions
Critical Analysis	Contextual and discursive interpretation of the data	Applying critical theory and discourse analysis to ethical language	Exposure of power dynamics, rhetorical strategies and institutional blind spots

Figure 2 Illustrating the structure or flow of data sources used in the study**.**

3.4. Analytical Framework

The research adopts thematic and discursive analytical methods. The research employs thematic analysis to locate common ethical problems, institutional reaction approaches, and structural conflicts that appear across different investigations. Close reading with iterative coding allows research to detect three main ethical motifs: fairness vs. efficiency, transparent system monitoring, and fragmented accountability in AI networks. The analysis reveals ethical subjects that result both from what the documents show and from the omissions within the documents. The author conducts a critical discourse analysis to analyze the rhetorical methods which frame ethical matters in the selected documents. The investigation explores language mechanisms through which authority gets established, legitimacy gets asserted, and risks get handled. The research explores technical terminology which prevents ethical problems from appearing, methods of procedural bureaucracy used for showing adherence, and institutional mechanisms through which ethical criticisms are made invisible. The research design combines analytic methods to detect explicit and implicit dimensions in the ethical involvement with AI.

3.5. Ethical Considerations and Reflexivity

The study acknowledges the ethical consequences which emerge from the nature of its research although it lacks traditional human involvement. The study investigates four vital matters which encompass systemic discrimination along with public surveillance and medical vulnerability as well as social justice while requiring substantial care and critical self-awareness and humility. Through ethical reflexivity researchers

structure their entire process to make decisions about selecting cases along with interpreting texts and building their arguments. The analyst maintains an active stance regarding positionality when researching since their academic background along with cultural and institutional factors determine their interpretive framework. The researcher utilizes this position to constantly analyze the representation power dynamics which affect subject groups who faced harm from AI systems. The research approach implements an intersectional analysis together with a decolonial perspective to give priority to excluded perspectives throughout both technological development and ethical studies.

3.6. Limitations of the Methodology

All research designs include specific shortcomings that should be considered. The analysis stops at publicly accessible materials because this exclusion prevents researchers from understanding private developer meetings, organizational demands, and proprietary algorithmic management decisions. The exclusive use of publicly available documents makes it difficult to collect authentic human perspectives regarding the interpretation of AI ethics. The author recognizes these limitations, as the methodology requires such trade-offs to achieve comprehensive depth, analytical coherence, and broad scope. While the qualitative interpretive framework has various associated limitations, it proves perfectly adequate for studying the present research topics. The research uncovers ethical AI as a discourse that changes based on its different meanings, rather than as static technical elements. Through its focus on discourse, institutional practice, and the political aspects of representation, this methodology creates an effective way to analyze potential ethical risks and opportunities of decision-making AI systems.

4. Result and discussion

4.1. Emergence of Ethical Narratives in Institutional Contexts

The analysis demonstrates how educational institutions, governmental bodies, and corporate entities use ethical narratives that balance innovation against responsibility when working with AI systems. Ethics functions throughout institutional documents as an opportunity which boosts technological advancement rather than restricting it. The terminology used to describe "trustworthy AI," "ethical innovation," and "responsible AI leadership" creates a link between ethical compliance and organizational market strength and public support. The apparent ethical alignment between ideals and operational needs leaves important discrepancies between principles and operational procedures unnoticed. Organizations use principles of fairness, transparency, and accountability in their statements, yet execute these principles in broad and selective ways, which leads to an ethical bare minimum stance that provides symbolic responsible management with limited real impact.

The assessment shows that ethical values have a central role in messaging, yet institutions leave behind effective enforcement tools for ethical principles. Ethical guidelines that organizations present to the public often lack compulsory force and effective oversight systems. The ethical engagement takes place through performative actions that substitute language for actual substantive conduct. Organizations use ethical audits and advisory boards as reputation-building tools instead of creating real ethical interventions. Self-regulatory methods in AI result in an ethical decision-making system that lacks institutional oversight, thus leading to uncertainty about both its effectiveness and purpose.

4.2. Conflicting Conceptions of Fairness and Bias

The treatment of fairness and bias in algorithmic systems emerges as the most significant ethical subject from the analysis. Every institution mentions fairness as an essential principle, yet they use different, inconsistent, and non-universal definitions in their institutional documentation. Different stakeholders implement fairness standards based on their organizational interests, which results in an unconnected ethical framework. The criminal justice system defines fairness through equal accuracy performance between population groups, but healthcare prioritizes equal availability of diagnostic resources among its patients. The multiple definitions of fairness exist beyond mere word choice because they stem from opposing moral values, statistical methods, and institutional demands.

The way bias receives treatment ignores its fundamental origins by focusing exclusively on mathematical aspects of technical bias while neglecting historical factors that cause inequality. Technical approaches to bias treatment exclude political considerations because they approach the issue as a solution-able engineering challenge instead of recognizing it as an outcome of structural discrimination. When fairness is designated as an algorithmic property instead of an institutional attribute, ethical accountability moves from people to machines. Organizations use such shifts to present technological solutions as progress toward better ethics, although substantive disparities continue to exist. Such ethical reductionism proves incapable of solving core moral dilemmas that arise from the use of AI decision systems in critical domains.

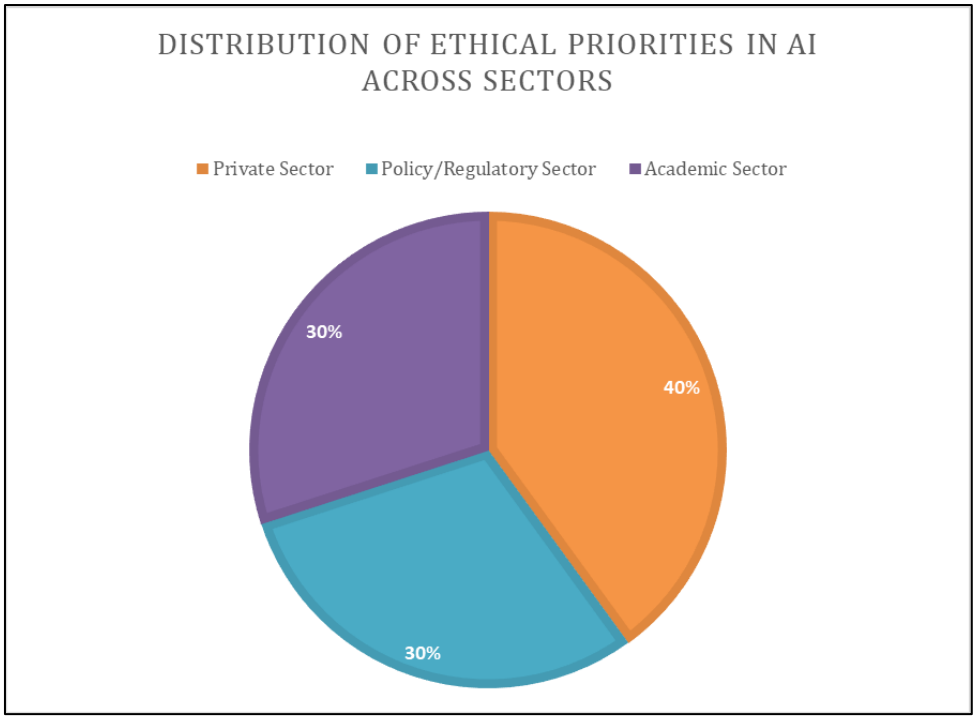

Figure 3 Distribution of Ethical Priority in AI Across Sector

4.3. Transparency as a Contested Value

Table 1 Interpretation and Challenges of Core Ethical Principles in AI Across Institutional Contexts

Ethical Principle	Common Institutional Interpretation	Operational Practice	Observed Challenges and Contradiction
Fairness	Equitable treatment across demographic groups	Bias mitigation via data balancing or model tweaking	Vague definition; lacks consensus; ignores structural inequality
Accountability	Assigning responsibility to system designers or regulators	Ethical audits, internal governance structures	Diffused responsibility; weak enforcement; limited transparency
Transparency	Disclosing system logic and decision pathways	Explainable AI (XAI), public statements	Often superficial; obscured by complexity and proprietary claims
Inclusiveness	Engaging "stakeholders" in development	Selective consultations, advisory panels	Excludes marginalized groups; engagement often symbolic
Privacy	Minimizing data misuse and surveillance	Data anonymization, encryption	Undermined by monetization imperatives and data dependency
Responsibility	Ethical awareness in design and deployment	Ethical review boards, guidelines	Lacks binding authority; ethics often performative or reactive

Ethical discussions about artificial intelligence feature transparency as their key concept, even as it serves as an all-encompassing solution to overcome public mistrust of AI systems and unclear decision algorithms. The study shows that transparency exists as a multifaceted principle, as different groups across institutions and sectors apply various definitions to it. Transparency covers different definitions depending on the setting where it is used, whether it requires technical documentation, model explainability, legal disclosure, or public consultation. Multiple interpretations of transparency create dysfunction between organizational implementation and evaluation.

Transparency becomes instrumental in practice, which generates opacity instead of clear visibility. The presented information typically represents narrow parts of whole pictures because entities behind these disclosures use strategic goals rather than ethical values. The technical details of information products, proprietary constraints, and organizational secrecy control meaningful information access for users. The current state of information openness shifts from its function as a democratic instrument into a superficial rhetorical approach. The design carries a perception of openness through its structure but maintains present-day control systems. Current configurations of transparency do not show sufficient evidence of tackling the epistemic inequalities which govern AI management.

4.4. Fragmentation of Accountability and the Ethics of Responsibility

One common pattern throughout the text addresses how AI accountability gets scattered across different steps from creation to implementation. Complex AI systems combined with many involved actors, including data providers, software engineers, system integrators, and end-users, produce fragmented or fully absent responsibility situations. When ethical issues arise, institutions build caseworks based on technological restrictions, regulatory holes, and undesired end results for avoiding responsibility and creating an impression of impartiality. Responsibility becomes impossible to enforce or attribute clearly because the concept spreads across several accountable parties.

An ethics of evasion appears when institutions admit their abstract responsibility but refuse to embed it with concrete institutional implementation. The evolution of technology happens more quickly than the legal sector develops new regulatory policies, and leaders of specific organizations commonly struggle to build or sustain adequate control systems. Ethical discourse takes the role of accountability placeholder within this lack of enforcement. The study demonstrates that to achieve genuine accountability, one must introduce systemic changes to governance systems through stakeholder integration and promote ethical criticism procedures within institutions. The absence of proper accountability systems makes the rhetoric of responsibility transform into an empty term that fails to stop or repair harmful outcomes.

4.5. Counter-Narratives and Critical Resistance

Research documents an expanding collection of counter-discourses which civil society organizations, investigative journalists, grassroots movements, and critical scholars develop from civil society. Critical narratives try to deconstruct main ethical guidelines by exposing gaps in their systems and showing their insufficient coverage. The interpretation sheds light on people who face adverse impacts from algorithmic systems, usually members of racialized, economically marginalized, or politically disenfranchised communities that official reports tend to overlook.

These viewpoints introduce new moral perspectives that draw from principles supporting justice, solidarity, and historical responsibility. These perspectives show how AI systems maintain the current social rankings and perform structural harm in the name of technical achievement and efficiency. These critiques shift the ethical analysis away from technical standards toward social effects, and from single human conduct toward organizational frameworks, to deliver an essential challenge to current ethical discussions. The critique seeks both ethical system design and democratic authority for managing AI development scenarios.

These patterns are further illustrated in the graph below, which visualizes the frequency with which ethical principles appear across corporate documents, policy papers, and academic articles.

4.6. The Political Economy of AI Ethics

Ethical discourse develops a strong connection with AI development's political economic dynamics as an important recurring motif in the reviewed documents. The institutions that develop ethical guidelines or conduct self-assessments present themselves as stakeholders who seek the wide acceptance of AI systems. The institutions serving in both ethical advisory and commercial sectors maintain an ongoing conflict between their public ethics role and their business strategic direction. The research reveals that ethical pledges, though presented with honest intentions, remain secondary to economic market objectives. Public scrutiny increases the likelihood of ethical initiatives' emergence, since organizations use ethics primarily for risk management instead of true ethical guidance.

The financial structures related to AI research and development control which ethical investigations receive focus and priority. Corporate academic research funding patterns direct what ethical questions get addressed in research because concerns must have technical solutions to pass through the funding process, despite ongoing system-level criticisms. Throughout supposedly non-biased policy sectors, an inherent bias emerges favoring economic theories that prioritize technological advancement and national competitiveness. This alignment creates an ethical discussion which prefers to examine future dangers and uncertain risks, although it disregards current injustices caused by AI-based systems.

Through this study, researchers conclude that the political economic framework of AI ethics requires ethical assessment because it determines the observations, actions, and communication which claim to promote ethical principles.

4.7. Institutional Memory and the Ethics of Forgetting

Research shows that institutions have weak capabilities for remembering past failures occurring in AI systems. Cases typically show a recurring pattern where previous problems, which included algorithmic discrimination, privacy breaches, and labor exploitation, are seen as unique incidents and not systematic dilemmas. The disappearance of public interest allows institutions to resume their previous procedures while making only small modifications to their processes. Academics have termed this ethical amnesia cycle as a barrier to both learning processes and the recognition of habituated harm within organizations.

By failing to deeply examine their issues, institutions reveal an ethical disregard for their responsibility to learn. The ability to forget ethical issues appears as one manifestation of a general societal practice which separates time from wrongdoings and wraps up stories to strip out harsh language. Framing in reports and corporate materials employs linguistic expressions such as "lessons learned" or "steps taken" to suggest ethical failures were properly addressed, thus blocking the chance for further examination. The strategy enables institutions to demonstrate change but prevent actual transformative initiatives. The analysis reveals that true ethical engagement necessitates institutions to develop a system for remembering patterns of harm and responsibilities through time by establishing enduring accountability systems. Institutions with no memory system maintain ethical performances instead of legitimate transformations, which yields shallow institutional improvements.

5. Conclusion

Research on ethical choices in artificial intelligence shows that principles, practices, and power systems connect, but do so in an uneasy manner. Multiple document research reveals ethics exists extensively in discussions yet remains minimally present in practice. The growth of Artificial Intelligence ethics as a rhetorical tool in governmental and corporate strategies and civil society advocacy yields only inconsistent real-world results because political and business aims dominate implementation.

Studies prove that stakeholders promote ethical ideas concerning fairness, accountability, transparency, and inclusion, yet rarely establish exact protocols or operational processes for enforcement. This disparity weakens the implications of noble ideas because essential principles lose their connection to actual power dynamics and social injuries. Institutional governance of AI uses prevailing ethical systems that focus on process protection measures and technological solutions but fail to analyze the fundamental social conditions affecting AI research processes and application contexts. This approach creates ethical minimalism which values public appearance over ethical justice and institutional status over ethical accountability.

What is most disturbing about the situation is that communities who face the highest impact from AI technology struggle to find appropriate representation. The existing ethical discussions mainly occur among privileged groups while disconnecting from practical realities, choosing to envision problematic future scenarios instead of acknowledging current adverse situations. Because of their exclusion from ethical discourse, these groups lose respect for ethical principles while their own problems remain unaddressed by artificial intelligence systems. When affected communities are left out of meaningful participation, their vital voices fail to enter ethical frameworks, which therefore remain controlled by members of the powerful elite.

The analysis strongly emphasizes the essential need to reconsider ethical conduct during contemporary AI development. Ethics must escape its current focus on principles and positioning by developing into a practice that depends on historical understanding, institutional responsibility, and collective resistance. The ethical framework needs to handle complicated situations while avoiding ambiguous boundaries and prioritize historically discriminated groups in technology governance. Ethics operates as a political venture beyond regulatory requirements or branding necessities, as it must distribute power fairly, make reparations for previous wrongs, and challenge technology's naturalistic visions.

The trajectory of AI development depends on human choices regarding the internal management of these systems, decisions about their beneficiaries, and protection of vulnerable groups from unwanted consequences. The quest for answers goes beyond code execution and regulatory compliance requirements because it requires critical thinking followed by brave actions for achieving justice across various disciplines. The promise of ethical AI will transform into actual practice only through ethical choices made by human beings.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2019 Jul;9(4):e1312.

[2] Nilsson NJ. The quest for artificial intelligence. Cambridge University Press; 2009 Oct 30.

[3] Benzinger L, Ursin F, Balke WT, Kacprowski T, Salloch S. Should Artificial Intelligence be used to support clinical ethical decision-making? A systematic review of reasons. BMC medical ethics. 2023 Jul 6;24(1):48. https://doi.org/10.1186/s12910-023-00929-6

[4] Rodgers W, Murray JM, Stefanidis A, Degbey WY, Tarba SY. An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. Human resource management review. 2023 Mar 1;33(1):100925. https://doi.org/10.1016/j.hrmr.2022.100925

[5] Etzioni A, Etzioni O. Incorporating ethics into artificial intelligence. The Journal of Ethics. 2017 Dec;21:403-18. https://doi.org/10.1007/s10892-017-9252-2

[6] Zhang Z, Chen Z, Xu L. Artificial intelligence and moral dilemmas: Perception of ethical decision-making in AI. Journal of Experimental Social Psychology. 2022 Jul 1;101:104327. https://doi.org/10.1016/j.jesp.2022.104327

[7] Yu H, Shen Z, Miao C, Leung C, Lesser VR, Yang Q. Building ethics into artificial intelligence. arXiv preprint arXiv:1812.02953. 2018 Dec 7. https://doi.org/10.48550/arXiv.1812.02953

[8] Guan H, Dong L, Zhao A. Ethical risk factors and mechanisms in artificial intelligence decision making. Behavioral Sciences. 2022 Sep 16;12(9):343. https://doi.org/10.3390/bs12090343

[9] Machado J, Sousa R, Peixoto H, Abelha A. Ethical Decision-Making in Artificial Intelligence: A

[10] Logic Programming Approach. AI. 2024 Dec 2;5(4):2707-24. https://doi.org/10.3390/ai5040130

[11] Benzinger L, Epping J, Ursin F, Salloch S. Artificial Intelligence to support ethical decision-making for incapacitated patients: A survey among German anesthesiologists and internists. BMC Medical Ethics. 2024 Jul 18;25(1):78. https://doi.org/10.1186/s12910-024-01079-z

[12] Bankins S. The ethical use of artificial intelligence in human resource management: a decision-making framework. Ethics and Information Technology. 2021 Dec;23(4):841-54. https://doi.org/10.1007/s10676-021-09619-6

[13] Ferrell OC, Harrison DE, Ferrell LK, Ajjan H, Hochstein BW. A theoretical framework to guide AI ethical decision making. AMS Review. 2024 Jun;14(1):53-67. https://doi.org/10.1007/s13162-024-00275-9

[14] Nunez C. Artificial intelligence and legal ethics: Whether AI lawyers can make ethical decisions. Tul. J. Tech. & Intell. Prop.. 2017;20:189.

[15] Huang C, Zhang Z, Mao B, Yao X. An overview of artificial intelligence ethics. IEEE Transactions on Artificial Intelligence. 2022 Jul 28;4(4):799-819. https://doi.org/10.1109/TAI.2022.3194503

[16] Abel D, MacGlashan J, Littman ML. Reinforcement Learning as a Framework for Ethical Decision Making. InAAAI workshop: AI, ethics, and society 2016 Mar (Vol. 16, No. 2).