ARTIFICIAL INTELLIGENCE ETHICS IN PHILOSOPHY AND HUMAN DECISION MAKING SYSTEMS

SRIKANTH V[1*], Anju Gautam[2], Amit Kumar Pandey[3], Dwijen Kirtania[4], Bhavna Pande[5], Kavya Singh[6], Seema Singh[7], Ajit Kumar[8], Raj Kumar Gupta[9]

[1]MCA Department, Acharya Bangalore B School, Andrahalli Main Road off Magadi Road, Bengaluru – 560091, India | Email: dr.srikanthv@abbs.edu.in

[2]Department of Computer Science and Engineering, Inderprastha Engineering College, Sahibabad, Ghaziabad (UP), India | Email: anjugautam330@gmail.com

[3]Department of CE (AI ML & DS), Marwadi University, Rajkot, Gujarat, India | Email: amitpy2@rediffmail.com

[4]Intuit inc, Adarsh Palm Retreat, Bellandur, Bengaluru, Karnataka 560103, India | Email: dwijen.kirtania@gmail.com

[5]School of Management, BBD University, Lucknow, India | Email: bhavnapande.bbd@gmail.com

[6]School of Management, BBD University, Lucknow, India | Email: singh.kavya08@gmail.com

[7]School of Management, BBD University, Lucknow, India | Email: meseemasingh@gmail.com

[8]Department of Business Administration, Hierank Business School, Ch. Charan Singh University, Meerut, India | Email: ajitmdh30@gmail.com

[9]AIML DEPARTMENT, JIMS Engineering Management Technical Campus, Greater Noida, India | Email: rajk428@gmail.com

***Corresponding Author:** SRIKANTH V | Email: dr.srikanthv@abbs.edu.in

ABSTRACT

Artificial Intelligence (AI) is undoubtedly one of the most game-changing technologies of the twenty-first century, affecting human decision-making in healthcare, finance, education, law and public governance. The potential of AI is immense, with benefits in efficiency, accuracy, and prediction but also with the growing number of ethical and philosophical questions. The study explores the ethical implications of AI in the context of philosophical frameworks and their implications for human decision-making processes. The study is qualitative and uses a systematic literature review and thematic analysis approach, which examines the most important ethical issues such as algorithmic bias, transparency, accountability, privacy, autonomy and moral responsibility. The results show that classical theories of ethics like utilitarianism, deontological ethics, and virtue ethics, and ethics of care are still very relevant for assessing today's AI systems. The research also found that while AI can improve decision-making accuracy and streamline processes, overreliance on AI-recommended decisions can lead to a reduction in human autonomy and accountability issues. In addition, AI systems are not, at present, considered to be an autonomous moral agent, which means that human oversight is necessary for good governance. The study identifies the key elements of a robust AI governance framework, such as embedding ethical considerations within regulatory frameworks, establishing accountability systems, and incorporating human-centralized design practices. These measures are essential to ensure that the development of technology continues to be in line with core human values, social justice and democratic principles and to build public confidence in decision making processes involving AI.

KEYWORDS

Artificial Intelligence, AI Ethics, Human Decision-Making, Algorithmic Bias, Moral Responsibility, AI Governance.

1. INTRODUCTION

1.1 Background of the Study

The advent of Artificial Intelligence (AI) in the computer science field has risen from theory to become an ever more influential technology in the way humans make decisions in various industries. The advent of recent machine learning, deep learning, NLP, and generative AI technologies has unlocked the potential of intelligent systems to execute complex tasks that once necessitated human perceptive abilities. Currently, AI technologies are being used in areas of health diagnostics, financial forecasting, educational systems, legal analysis and public governance, impacting decision making at a societal and individual level (Sun et al., 2024).

As AI's adoption has progressed, so has the use of decision-making systems that rely on automation. AI algorithms are employed to assess credit risk, foresee criminal activity, help in medical diagnosis, guide educational planning, and enhance the effectiveness of public service provision in organizations. These systems offer improved

efficiency, precision, and scalability over traditional human-driven systems. With the increasing autonomy of AI systems in decision-making, however, questions about the ethical implications and societal impacts arise (Cheong et al., 2024).

As AI infiltrates more and more critical areas, questions on fairness, accountability, transparency and human oversight have become more prominent. Human-centred AI systems have been highlighted by international bodies like UNESCO and the Organisation for Economic Co-operation and Development (OECD) as systems which should support human rights, democracy and social justice (UNESCO, 2024; OECD, 2024). As AI becomes more influential in human decisions and actions, it sparks profound ethical, moral, and legal dilemmas about its impact on human autonomy and accountability. As AI gains greater influence in human judgment and action, it poses essential philosophical questions of morality, responsibility, and maintaining human autonomy and accountability. Ethical evaluation has therefore become imperative, making sure that technological development is in line with the values and ethics of society.

1.2 Problem Statement

While AI-powered decision-making tools provide great promise of benefits, many ethical issues have not been addressed. The inability to understand how decisions are made, known as the 'black-box' problem, is one of the main concerns. This can make it challenging to take responsibility and accountability when AI systems generate negative and discriminatory results (Cheong et al., 2024).

Additionally, algorithmic bias has emerged as a major ethical issue. Artificial intelligence systems that are trained with biased data can exacerbate social inequalities and discrimination against certain groups in society. The use of AI in governmental and organizational decision-making has been a cause for concern, with recent investigations revealing instances of unfairness towards persons due to their age, nationality, and disability, among other factors (The Guardian, 2024). In addition to technical issues, there is a philosophical question of whether machines can make ethically driven decisions, or simply mimic patterns in the data. The more AI is used to make decisions for people, the more there are concerns that it may be taking away the human element, moral judgment, and personal responsibility from decision-making. Such challenges require a thorough ethical and philosophical analysis of decision-making systems that are based on AI.

1.3 Aim and Objective of the Study

The main goal of this research is to consider the ethical aspects of Artificial Intelligence in the context of philosophical discussions and in human decision-making processes. This study will seek to answer the question of how to:

- Discuss the philosophical principles underlying AI ethics.
- Discuss ethical issues with AI-based decision-making systems.
- Critically assess the implications of using AI technologies on autonomy, responsibility and moral judgment of humans.
- Discuss topics of transparency, accountability, fairness and bias in algorithmic systems.
- Suggest ethical frameworks and governance strategies that enable ethical AI development and use.

1.4 Research Questions

The following research questions will be addressed in this study:

1. Which Philosophical theories and ethical principles are most relevant to evaluate the Artificial Intelligence?
2. How can AI affect the human decision-making systems in various fields?
3. What are some ethical issues of using algorithmic decision-making systems?
4. What are the implications of transparency, accountability and bias in AI systems?
5. What opportunities exist for the effective application of ethical governance frameworks to complement accountability and responsible actions on AI implementation?

1.5 Significance of the Study

This research work is part of an increasing number of studies in the field of Artificially Intelligent and Philosophy / Ethics. Academically, it helps to promote the understanding of the application of classical ethical philosophy to the contemporary challenge of technology. The study also gives a philosophical view on the issues of autonomy, responsibility, and moral agency in AI mediated environments.

The study's results could also help policymakers, regulators, technology builders, and organizational leaders create AI systems that are both fair and transparent and promote human well-being. The value of ethical guidance based on philosophical analysis grows as governments and international organizations work on establishing frameworks for the governance of AI. Overall, the project contributes to the building of human-centred AI systems that foster innovation and preserve fundamental rights, social justice and democratic values (OECD, 2024; UNESCO, 2024).

2. LITERATURE REVIEW

2.1 Understanding what Artificial Intelligence is and the ethical considerations

Artificial Intelligence (AI), in itself, is an umbrella term for a group of computational systems that can carry out tasks that usually require human intelligence, such as learning, reasoning, perception, language processing, and decision making. (Sun et al., 2024) Generally, AI can be divided into narrow AI that can only perform specific

tasks and more advanced forms of AI that show more cognitive abilities. Advancements in machine learning and generative AI have greatly broadened the scope of practical applications for intelligent systems in sectors and their impact on daily life and decision-making processes in institutions (Sun et al., 2024).

Given the increasing use of AI, UNESCO (2024) states that there are pressing ethical questions about human rights, privacy, fairness, accountability, and social justice. The ethical challenges that have been discussed in the past are no longer only technical, but also questions of society as a whole regarding the design, governance and assessment of intelligent systems. With AI technologies increasingly being integrated into healthcare, education, finance, and governance, promoting ethical practices of AI systems is a global priority (UNESCO, 2024). Ethical issues are especially important when AI systems impact opportunities, freedoms, and individual well-being.

2.2 The Concept of Ethics in Philosophy

Takeshita et al (2023) note that utilitarian ethics, first formulated by Jeremy Bentham and John Stuart Mill, is still very relevant to AI ethics because it focuses on the outcomes of an action and the general social good. Most AI systems are engineered to make things work most efficiently, most productively, most well. Most AI systems are built in a utilitarian way. But critics say a bias toward outcomes can neglect individual rights and minority concerns when the decision of the algorithm puts an emphasis on the aggregate benefits (Takeshita et al., 2023).

AI governance discussions are still highly relevant to deontological ethics, especially the philosophy of Immanuel Kant, as noted by Radanliev (2025). The Kantian ethics emphasize duties and moral obligations, and respect for human dignity, not only consequences. This view should never consider humans as data resources or efficiencies for AI systems. Human rights, autonomy, and informed consent are key ethical principles that transcend the benefits of technology (Radanliev, 2025).

In Aristotle's virtue ethics, they promote the development of moral character, wisdom, and responsible judgment, principles that have been applied in the field of AI development to foster more responsible and ethical use of the technology. Virtue ethics, a principle from Aristotle, has been linked to the development of AI, with a focus on promoting moral behaviour, wisdom, and responsible judgment in the use of the technology. Instead of wondering if actions with AI will be most useful or meet rules, virtue ethics asks whether technology will foster human flourishing and ethical action. This method pushes developers and institutions to think about greater society's values in the design of intelligent systems (Brand & Nannini, 2023).

UNESCO (2024) also highlights the relationships, empathy, social responsibility and caring for vulnerable groups as the central tenets of contemporary ethical approaches such as ethics of care. They also build on existing ethical frameworks and take account of the social context in which an AI system is used, which means that fairness, inclusion and welfare must be salient dimensions of all ethical considerations (UNESCO, 2024).

2.3 AI and Moral Agency

Zafar (2025) explores the relatedness of moral agency in AI and discusses the “debate” of whether intelligent systems can actually have moral agency. The use of advanced AI systems can lead to systems acting autonomously, but it does not necessarily mean that they have consciousness, intent, or moral understanding like a human. However, it is still a topic for debate whether machines have moral responsibility or not (Zafar, 2025).

In contrast, Fritz (2020) argues that computational systems have to be separated from human moral agents because responsibility is ultimately the responsibility of the humans that design, deploy and supervise technological systems. While algorithms might play a role in the outcomes generated by AI, machines cannot be held completely responsible for ethical responsibility. This still matters in comprehending responsibility in human-AI interactions (Fritz, 2020).

This means that AI must be considered by organizations within the context of organizational decision-making, not as a moral entity in its own right, as Schmitz and Bryson (2025) suggest. They put a strong emphasis on human accountability, supervision and avoiding "diffused responsibility" that can happen when too much trust in automation by institutions. This kind of approach will enhance transparency and ethical governance and maintain the human element in decisions with consequences.

2.4 Human Decision-Making Systems

In the past, traditional theories of decision making have tended to focus on the idea that people make decisions in a rational manner, by weighing the information they have available and choosing the option they believe will lead them to the highest possible return. The rational decision-making models remain relevant in economics, management, and policy analyses due to the structured nature of their problem-solving and choice evaluation process (Cheong et al., 2024).

Manoli et al. (2024) have pointed out that the belief in a complete rational model was challenged by behavioural studies which have shown that human decisions often rely on heuristics, cognitive shortcuts and emotions. When the situation is uncertain, complex, or information overload, people might use simple decision-making processes and make decisions that do not conform to the ideal rational standards. The results of these investigations are relevant to the understanding of human-AI interaction in terms of recommendations.

In addition, cognitive biases like confirmation bias, anchoring bias, and overconfidence can have a substantial impact on human decision-making (Brand and Nannini, 2023). With artificial intelligence systems becoming more

involved in decision-making, it is crucial to understand these restrictions as technology can either help to reduce or amplify these biases in design and application.

2.5 The impact of AI on decision-making

According to Google AI (2025), AI-driven decision support systems are becoming more common for analysing information, spotting patterns, and providing recommendations for people and organizations. These systems help to optimize processes by analysing massive amounts of information that a human brain could not handle on its own, thus supporting quicker and more well-informed choices in different industries.

According to Brand and Nannini (2023), AI recommendations have the potential to affect human behaviour by impacting perceptions, preferences, and judgements. Explainable AI tools can enhance human understanding and aid in making informed decisions, while opaque tools can foster a culture of blind trust in the algorithms. This is a topic of ethical concern about autonomy and informed consent.

However, Cheong et al. (2024) warn that overreliance on AI suggestions can lead to a type of algorithmic dependency, where people increasingly rely on technology for decision-making. The reliance on such systems can diminish the ability for critical thinking and diminish the responsibility of humans in decision-making processes, especially since AI-generated content might seem objective or authoritative, yet may hold limitations or biases (Cheong et al., 2024).

2.6 Learning and teaching of algorithmic bias and fairness

Algorithmic bias is one of the most critical ethical issues with AI systems, according to Radanliev (2025). Bias can arise from unrepresentative data, faulty assumptions about the design, historical inequities, or from the contexts in which the tools are used. These biases can be embedded in algorithms and create discriminatory results for individuals and communities.

Automated systems used in public sector decision making are also subject to evidence of bias, as seen in the Guardian (2024).There is evidence of bias in automated systems used in public sector decision making, as reported in the Guardian (2024), and the disproportionate impact such systems can have on vulnerable populations. These revelations highlight how AI systems can be socially biased and perpetuate social inequities without proper ethical safeguards. These concerns are especially pertinent in the fields of employment screening, financial services, healthcare diagnostics, and criminal justice.

According to UNESCO (2024), fairness and non-discrimination must be key principles when it comes to AI governance. Bias should be detected, tracked, and addressed across the whole system lifecycle in ethical AI systems. Fairness is a process of continuous evaluation, engagement with stakeholders, and mechanisms of accountability to ensure that harms that may occur are taken into account before they become entrenched.

2.7 Tools and techniques to increase transparency, explainability and accountability

According to Brand and Nannini (2023), transparency and explainability are two of the key ethical demands due to the fact that many AI systems are black-box algorithms, whose internal decision-making process is not easy to understand. If there are no explanations, the people concerned might lose ability to question decisions or even evaluate if outcomes are fair and warranted. Explainability is thus a means of supporting trust and accountability.

As the number of autonomous systems deployed continues to grow, the need for explainable AI is gaining significance, according to TechRadar (2026). Transparent models make it easier for users to grasp how decisions are produced, can be audited, and can be used to enhance the compliance aspect with the regulations. Explainability can also be used to detect errors, biases, and unintended consequences before they cause a meaningful harm.

Schmitz and Bryson (2025) state that accountability becomes an issue when responsibility is spread across developers, organizations, users and the AI systems themselves. Good governance involves clear responsibility, human monitoring and control of the performance of the system. These measures play a crucial role in maintaining human responsibility for the actions resulting from AI-driven decision-making.

2.8 Matrix of existing ethical frameworks for AI governance

The AI Act is the EU's first comprehensive legal framework dedicated to the regulation of AI, introduced by the European Commission (2024). The bill takes a risk-based approach that has more stringent obligations for high-risk uses of AI and encourages innovation and trust in AI. The AI Act is a major step toward enacting legally binding requirements for the responsible development and use of AI.

OECD (2024) recommends AI Principles rooted in inclusive growth, values grounded in human rights, transparency, robustness, accountability and respect for democratic institutions. These principles offer practical guidance to governments and organizations to create responsible, trustworthy AI systems while fostering innovation in AI.

UNESCO (2024) seeks to promote a global approach by publishing its Recommendation on the Ethics of Artificial Intelligence, which emphasizes human rights, fairness, transparency, sustainability and international cooperation. The framework promotes ethical factors at every stage of the AI lifecycle and promotes meaningful oversight by humans.

The United Nations Advisory Body (2024) points out the increasing overlap in the international governance agendas. While there are disparities in the use of enforcement mechanisms and the nature of regulation, there are

shared principles that have been captured in most frameworks such as transparency, accountability, fairness, human rights protection and human-centred governance. The convergence of this suggests a global consensus of ethics, which could lead to the formulation of more future AI regulations and policies.

3. RESEARCH METHODOLOGY

3.1 Research Design

This study is qualitative in nature with conceptual and analytical inquiry. The aim of the research is not to determine statistical relations, but to critically analyse the ethical aspects of Artificial Intelligence (AI) in philosophy and human decision making systems. Given the nature of ethical issues related to AI that include abstract concepts like autonomy, accountability, fairness, transparency, and moral responsibility, which cannot be quantified, a qualitative approach is well-suited. The study aims at elucidating the implications of ethical theories for modern AI systems and to revealing the effect of AI technologies on human decision-making processes through philosophical analysis and critical reading of the literature. This method has been utilized extensively in the study of AI ethics, as it can facilitate the examination of intricate ethical dilemmas and governance issues (Batool et al., 2024; UNESCO, 2024).

3.2 Data Sources

This study uses only secondary data, which is collected from scholarly and institutional sources. The main sources of information are academic journal articles, books, conference proceedings, policy reports, and documents from international governance. Specific attention is given to publications published since 2020 and up to 2025 to ensure the analysis' relevance and currency. Institutional sources include the UNESCO Recommendation on the Ethics of Artificial Intelligence, OECD AI Principles, European Union AI Act documentation and reports from international organizations related to AI governance. These sources serve as a valuable resource for understanding the philosophical discussions, regulatory implications and ethical considerations of AI technologies. Secondary data is deemed as appropriate since the study is not designed to produce fresh empirical data, but to compile available knowledge..

3.3 Data Collection Methods

The data were gathered by a systematic literature and policy analysis of relevant sources. The literature review was done by searching in the peer-reviewed literature covering the aspect of AI ethics, algorithmic decision making, transparency, accountability, fairness, and the human autonomy. Publication relevance, methodological rigor, academic credibility and recency were used as selection criteria. The international policy frameworks and governance guidelines were also looked at to offer real-world insights into the use of ethical AI. The document analysis technique helped the researcher to detect the themes, ethical issues and governance that were repeated throughout various sources. This method is frequently employed in AI ethics studies due to its ability to allow for extensive discussions of the normative aspects, institutional reactions, and the evolving nature of regulation (Khan et al., 2021; Batool et al., 2024).

3.4 Analytical Framework

Thematic analysis and philosophical analysis are used to analyse the study. Themes and patterns in the literature on AI ethics and algorithmic governance, human autonomy, accountability, fairness, transparency, and moral responsibility are explored through thematic analysis. Academic research results, governance structures and policy statements are sorted into thematic areas according to the research objectives.

Using pre-existing ethical frameworks, such as utilitarianism, deontological ethics, virtue ethics and ethics of care, philosophical analysis is used to assess ethical issues surrounding the use of AI. This framework provides a systematic way to explore the opportunities and risks of AI decision-making systems from various perspectives. Comparative analysis is also used to compare and contrast the key international governance instruments such as the UNESCO Recommendation on the Ethics of Artificial Intelligence, the OECD AI Principles and the European Union AI Act. The study aims to provide a holistic understanding of ethical governance of AI through this multi-faceted analysis and to pinpoint practical measures that can guide the integration of technology with human values and social responsibility. The current body of literature on AI governance puts a strong focus on the need for a combination of normative principles and institutional governance and accountability mechanisms, and on the importance of embedding these in a lifecycle approach (Mäntymäki et al., 2022; OECD, 2024).

3.5 Ethical Considerations

This study is not a study of people, but ethical considerations are still important while undertaking a research project. The researcher demonstrates academic integrity by citing and paraphrasing evidence accurately and interpreting the evidence objectively, and by responsibly using scholarly sources. The aim of avoiding bias is achieved by including different points of view across the disciplines of philosophy, computer science, public policy and governance studies. In addition, the study is conducted in a transparent and honest approach, with clear limitations identified as caveats to secondary data analysis. AI ethics is a normatively driven field, so it is crucial to ensure analytical neutrality to assess differences in ethical perspectives and governance frameworks (UNESCO, 2024; OECD, 2024).

4. RESULTS AND DISCUSSION

4.1 AI Ethics from a Philosophical Perspective

The literature review sheds light on the ethical issues related to AI that have long existed in classical philosophy. The review of the literature reveals the philosophical roots of the ethical implications of AI. AI is a new technology, but many of the ethical issues that are raised will be philosophical in nature: What is moral, what is just, who is responsible and what about human dignity? Recent research suggests that utilitarianism, deontological ethics, virtue ethics, and ethics of care remain significant approaches to considering ethical issues with AI systems and their social impacts (Sun et al., 2024; UNESCO, 2024).

Utilitarianism evaluates AI systems based on their impact and utility in society. In this context, AI is viewed as a morally beneficial tool because it can help to increase efficiency, minimize mistakes, and have a positive impact on health and societal well-being. For instance, AI-powered medical diagnosis has shown remarkable promise in enhancing the efficiency of disease detection and treatment planning. In a similar fashion, AI-driven public services can enhance resource management and boost administrative effectiveness. But utilitarian arguments can lead to decisions that benefit the majority as a group even if the harms are suffered by minorities, raising concerns of fairness and distributive justice (Takeshita et al., 2023).

Deontological ethics is the other viewpoint which focuses on duties, rights and respect for human dignity. Whether an AI system is more efficient or not, it should still respect fundamental human rights when making decisions involving employment, healthcare access or legal outcomes. A Kantian view would be that people should never be used as data points or statistical categories. This method emphasizes the potential threats to privacy, monitoring systems, and automated decision-making, which could compromise personal agency (Radanliev, 2025).

The concept of virtue ethics redirects focus from the capabilities of AI towards the moral character and intentions of its creators, builders, and overseers. Virtue ethics not only considers how AI impacts outcomes and rules, but also whether it promotes human flourishing and societal well-being. Responsibility, prudence, transparency, and fairness of AI developers and policymakers are thus necessary virtues that need to be developed to create ethical AI products (Brand & Nannini, 2023).

Table 4.1 Philosophical Perspectives Applied to AI Ethics

Ethical Theory	Core Principle	Application to AI
Utilitarianism	Maximizing overall welfare	AI improves efficiency and societal outcomes
Deontology	Respect for rights and duties	Protection of privacy, autonomy, and dignity
Virtue Ethics	Promotion of moral character	Responsible AI development and governance
Ethics of Care	Focus on relationships and vulnerability	Inclusive and human-centred AI systems

Table 4.1 displays the different criteria for assessing AI systems among the different philosophical frameworks. The various ethical viewpoints indicate that there is no single comprehensive theory that can solve all the ethical issues raised by the use of AI.

Figure 4. 1 Empty conceptual connection between AI Governance and two major ethical theories.

In general, the classical ethical theories help influence contemporary rules of governing AI, as shown in figure 4.1. These theories show the multi-dimensional nature of AI ethics, which involves both technical and philosophical aspects.

4.2 AI and Human Autonomy

The influence of AI on human autonomy is one of the most crucial aspects that have been revealed by the literature. Autonomy is the ability of a person to make decisions for themselves, independently, based on their own thinking and sense of right and wrong. With the growing presence of AI systems in decision-making contexts, there has been growing worry about how much influence, shape, and/or potentially undermine they have on autonomous human decisions (Cheong et al., 2024).

Recommendation systems are used more and more by AI to influence consumer behaviour, educational decisions, healthcare decisions, and political involvement. These systems can offer useful information and recommendations; however, they could also affect decisions in ways that could not be understood by users. Algorithms filter, show,

and recommend certain information to users and filter and prioritize alternatives. Human decision-making, therefore, can become algorithmic, instead of rational.

There are also recent studies that indicate that overreliance of AI recommendations can lead to "automation bias" as people will trust the AI output without taking note of possible errors or limitations. In high-stakes settings like healthcare or legal proceedings, human professionals could rely on AI-generated suggestions even if there is conflicting data available (Manoli et al., 2024).

Table 4.2 Effects of AI on Human Autonomy

Positive Effects	Negative Effects
Faster access to information	Automation bias
Enhanced decision support	Reduced critical thinking
Increased efficiency	Dependence on algorithmic recommendations
Improved analytical capabilities	Potential erosion of human judgment

AI can both increase and decrease autonomy, as illustrated in Table 4.2. The question is how to provide the technologic support without losing sight of the need for good human direction.

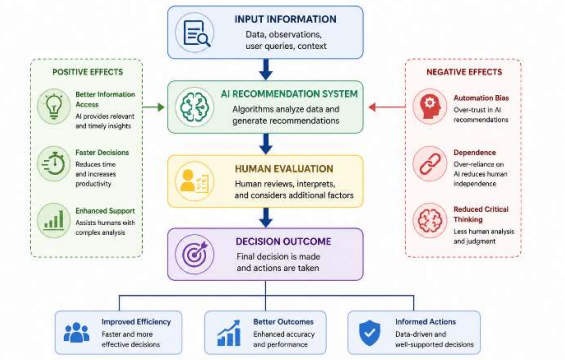

Figure 4. 2 The correlation between algorithmic recommendations, human judgment and decision outcomes

The correlation between algorithmic recommendations, human judgment and decision outcomes is illustrated in Figure 4.2. The figure underscores that if proper safeguards are not in place, there may be a diminished ability to make independent decisions as algorithmic influence rises.

The other key finding relates to the difference between "assisted decision-making" and "delivered decision-making. Assisted decision making means that AI gives recommendations and humans have final say. Delegated decision-making is when decisions are delegated successfully to automated systems. Ethical issues may be intensified when delegation is used in place of real human involvement, leading to a loss of accountability and moral responsibility (OECD, 2024).

The entire literature makes a case for a human centred approach, where AI is used not instead of, but in addition to human judgment. This allows for maintaining autonomy and taking advantage of technology capabilities. Even with the possibilities of AI, human oversight is still required as moral reasoning requires contextual understanding, empathy and value judgement that current AI cannot fully replicate (UNESCO, 2024).

4.3 Ethical Issues in Artificial Intelligence Decision Making Systems

Ethical issues emerged repeatedly in the thematic analysis, covering several aspects of AI's impact on making decisions. Of those, algorithmic bias, transparency problems, privacy concerns, accountability problems and social inequality were among those issues that were most commonly discussed in the current literature.

One of the biggest ethical issues is algorithmic bias. Any social bias in the historical data used for training AI systems can be perpetuated or intensified by the algorithms or models they create. They've been observed in hiring software, credit scoring, medical diagnostics, and predictive policing. This can manifest in the form of discrimination against people based on their race, gender, age, socioeconomic status, or other factors (Radanliev, 2025).

Table 4.3 Major Ethical Challenges in AI Decision-Making Systems

Ethical Challenge	Description	Potential Impact
Algorithmic Bias	Biased training data and models	Discrimination and inequality
Transparency Deficits	Lack of explainability	Reduced trust and accountability
Privacy Concerns	Extensive data collection	Violation of individual rights
Accountability Gaps	Unclear responsibility	Difficulty assigning blame
Social Inequality	Unequal technological effects	Marginalization of vulnerable groups

As shown in Table 4.3, ethical issues go beyond the technical performance and have a social and moral dimension. The issue of transparency is also crucial. Many AI systems are "black boxes" in which users and regulators, as well as developers, struggle to comprehend the decision-making processes used. If they can't see how the decision

was reached, people might lose their faith in the AI system. Transparency isn't always simpler to discover errors, dispute unfair experiences, or guarantee adherence to ethical criteria (Brand & Nannini, 2023).

As a result of the proliferation of large-scale data collection and surveillance technologies, privacy concerns have increased. AI systems can need a lot of personal data in order to work properly. This information can enhance the predictive value, but also poses questions about informed consent, data security, and personal rights. The unauthorized use or misuse of personal data can have serious ethical and legal implications (UNESCO, 2024).

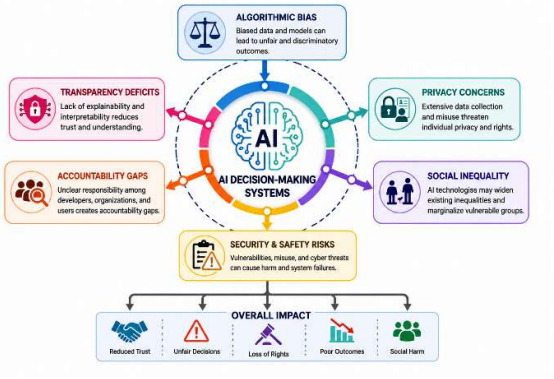

Figure 4. 3 Ethical Challenges Associated with AI Decision-Making Systems

Thematic analysis and literature review identified the major ethical challenges and the results are mapped on the figure 4.3. The illustration above shows the interrelatedness of the concerns of bias, transparency, and privacy, and accountability.

Another obstacle is accountability. The attribution of blame to AI systems becomes more intricate when they lead to negative consequences. They can be shared with the developers, organizations, users and policy makers. This diffusion of responsibility can lead to gaps in accountability if there is no one who is solely responsible for the negative impacts (Schmitz & Bryson, 2025).

The other major finding is related to the social impact of inequality resulting from AI. AI technologies might exacerbate the technological divide, favouring advanced groups and institutions over the less advanced ones. If-designed inequity or exclusion is possible when AI is not intentionally designed to address inequity or inclusion.

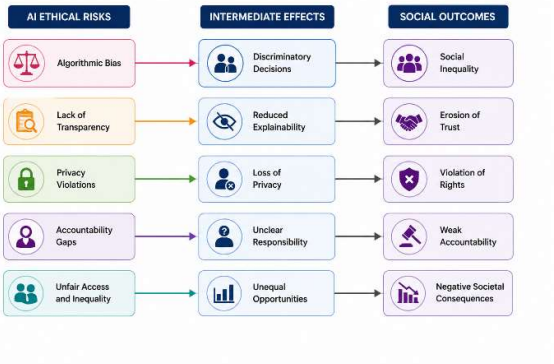

Figure 4. 4 Framework Linking AI Ethical Risks to Social Outcomes

A conceptual model is shown in Figure 4.4, illustrating the relationship between technological risks and wider societal impacts. The framework illustrates the possible link between algorithmic ethics breach and discrimination, loss of trust, loss of autonomy, and social inequality.

Overall, the results suggest that it is not possible to solve ethical issues in AI decision-making systems with a technical approach alone. Collaborations of multiple disciplines - philosophers, computer scientists, civil society organizations, and policy makers - are essential for effective solutions. The evidence indicates that transparency, accountability, fairness, human oversight and rights based protections need to be embedded in the AI lifecycle. This is not only aligned with the developing frameworks of international governance, but it also promotes the creation of trusted, human-centric AI systems that can provide societal benefits, with ethical risks reduced.

4.4 Impact on Professional Decision-Making

The thematic analysis shows how the use of AI has revolutionized professional decision making in many fields, including healthcare, finance, education, law and public administration. AI systems are more widely used as decision support systems, helping experts to analyse vast amounts of data, discover patterns, and make suggestions. All these capabilities have brought a greater efficiency and accuracy to the field of practice, but have also raised critical ethical issues related to professional judgment and accountability.

In the medical field, AI-powered diagnostic tools have shown promise in the diagnosis of diseases and making clinical decisions. Medical imaging, patient histories, and diagnostic records can all be analysed by machine learning algorithms, which can process these records more quickly than a human. This has led to the use of AI-driven recommendations gaining traction as a valuable tool for healthcare professionals in treatment decisions. Despite these systems' potential to enhance clinical outcomes, there are concerns about potential algorithmic bias and the risk of algorithmic recommendations being overly prescriptive, leading to a potential reduction in clinical judgment. Thus, human presence is crucial to ensure professional responsibility and patient-centred care (UNESCO, 2024).

Another sector that is profoundly impacted by AI is finance.AI is also very influential in the financial sector. AI is used in financial institutions for credit scoring, fraud detection, risk management, and investment predictions. Such applications help to enhance the efficiency of operations and processing time. But the results suggest that algorithmic decisions could inadvertently perpetuate biases in past financial data. Therefore, if fairness protections are not properly put in place, people might be facing unequal access to finance. The ethical dilemma is the need to trade-off between the predictiveness of the model and fairness, transparency, and non-discrimination (Radanliev, 2025).

AI-powered learning systems are being used more and more in educational institutions to tailor instruction to an individual student and to evaluate student performance and suggest educational paths. Personalized learning has the potential to enhance learning outcomes however, there have been issues raised about privacy of data, algorithmic profiling, and loss of teacher autonomy. Overreliance on the automated suggestions for learning can reduce the way the student learns the whole story from the human educators when assessing student needs and potential.

AI has also been integrated into the legal and judicial sectors to assist in case analysis, document review, and risk assessment. In some jurisdictions, recidivism risk is assessed using predictive algorithms to help with sentencing. For instance, these tools can boost consistency and productivity, but critics say algorithmic predictions can exacerbate systemic inequalities if they are based on historical inequalities in the data. Ethical issues are of high importance in relation to fundamental rights and freedoms making a difference in judicial decision making (UNESCO, 2024).

Table 4.4 Impact of AI on Professional Decision-Making Across Sectors

Sector	AI Application	Benefits	Ethical Concerns
Healthcare	Diagnostics and treatment support	Improved accuracy and efficiency	Automation bias, patient autonomy
Finance	Credit scoring and risk analysis	Faster decision-making	Algorithmic discrimination
Education	Personalized learning systems	Customized instruction	Data privacy and profiling
Law	Predictive analytics and case review	Consistency and efficiency	Bias and accountability
Public Governance	Resource allocation and service delivery	Operational effectiveness	Transparency and fairness

As shown in Table 4.4, AI applications can have significant benefits but also pose ethical risks that must be monitored. The results indicate that expertise of professionals should be leveraged in addition to algorithmic systems. Current AI technologies are unable to replace human judgment in many professional decisions, as they lack the ability to understand the context, show empathy, or engage in moral reasoning.

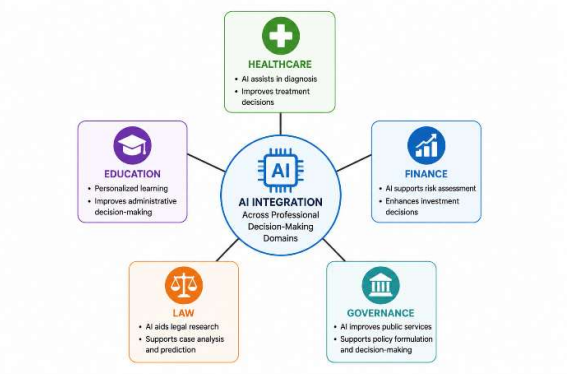

Figure 4. 5 AI in Professional Decision-Making Environments.

The figure depicts the increasing penetration of artificial intelligence technologies in various professional areas, while also revealing the trade-offs between AI benefits and ethical considerations.

4.5 Human Responsibility versus Machine Responsibility

One of the most important philosophical challenges found in this research relates to the distribution of responsibility's role in the context of AI support in decision-making processes. With the increased autonomy and recommendation-making power of AI systems, a dilemma has emerged: who will be held accountable if there is a negative impact?

The results show that it is generally not a position that is taken for granted by contemporary philosophical and legal views that an AI system itself is morally responsible. While AI can undertake complex cognitive functions, it has no sense, intent, morality or true autonomy in the philosophical sense. Therefore, AI systems are not beheld as moral agents in the same way as humans. The responsibility must thus lie with the individual and institutional entities that design, implement and govern these technologies (Schmitz & Bryson, 2025).

Developers are central to their work as they are responsible for defining the algorithmic architecture, training set, system goals, and operating limitations. The behaviours and ethical consequences of a system can be greatly affected by the design of that system. For that reason, it is the developer's responsibility to ensure that their AI system is in accordance with the ethics of fairness, transparency, accountability, and non-discrimination.

Businesses operating AI systems have extensive accountability as well. The integration, monitoring and implementation of the use of AI technologies are a matter for institutions. Lack of adequate control arrangements can lead to adverse results irrespective of the successful operation of the technical systems. Thus, organizational accountability goes further than technical performance to governance mechanisms, risk management mechanisms, to ethical review mechanisms (Mäntymäki et al., 2022).

Similarly, users and decision makers have significant roles to play. Do not blindly follow the suggestions of AI. Human operators are still responsible for assessing outputs, taking into account contextual factors, and making the final judgement on decisions. Meaningful human oversight is more and more a part of ethical governance frameworks, as it is simply not possible to fully delegate human accountability to machines (OECD, 2024).

Table 4.5 Responsibility Distribution in AI Decision-Making Systems

Stakeholder	**Primary Responsibilities**
AI Developers	Ethical design, testing, bias mitigation
Organizations	Governance, monitoring, compliance
End Users	Critical evaluation and responsible use
Regulators	Standards, oversight, enforcement
Society	Democratic participation and ethical scrutiny

As shown in Table 4.5, AI outcomes are shared among various stakeholders. This shared-responsibility approach takes into account the fact that ethical risks do not arise from a single point in the deployment of AI, but continuously throughout the AI lifecycle.

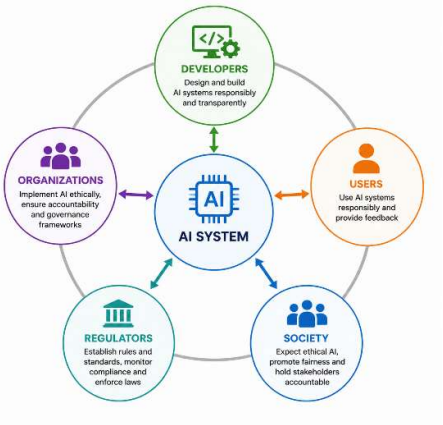

Figure 4. 6 Multi-Stakeholder Accountability Framework for AI Systems

The relationships between developers, organizations, regulators and users through a shared accountability framework are shown in figure 4.6.

The analysis also shows that as AI systems become more complex and autonomous, the accountability process becomes more challenging. Where the bad thing or result cannot be easily connected with one specific actor, there is a "responsibility gap" problem. Such gaps are linked to a loss of public confidence and to a reduction in the effectiveness of government. Such gaps are said to erode the effectiveness of government and public trust. Hence, traceability, audit trails, documentation regulations and explainability requirements have become integral features of modern day AI governance frameworks (OECD, 2024).

The European Union AI Act and new international governance frameworks have been keenly discussed in recent years, highlighting the need to ensure that clear accountability frameworks are preserved. Such frameworks aim to hold human decision-makers accountable for actions taken with AI technologies, even if they are heavily used

in the decision-making process. These help to maintain public confidence and avoid confusion about responsibility due to technology.

4.6 Ethical Governance of AI in the future

Finally, the analysis revealed a theme around the future trajectory of ethical governance of AI. As AI tools and capabilities grow, the importance of addressing ethical issues has become evident. Good governance needs to be supported by a robust framework to implement ethical principles in terms of organizational and regulatory processes.

A key finding is the growing convergence of the international governance frameworks on common principles. Transparency, accountability, fairness, human rights and meaningful human oversight are key principles in creating trustworthy Artificial Intelligence, as outlined in the UNESCO Recommendation on the Ethics of Artificial Intelligence, OECD AI Principles and the European Union AI Act. This alignment implies that there is a global "consensus of ethics" over the development of responsible AI.

But the application of ethical principles in practice is a big challenges. There are various groups that promote ethical principles for AI, but lack actionable governance. Lifecycle considerations are noted both for the design and development of AI systems and for their deployment, monitoring and retirement, with ethical issues starting from the design phase and continuing through the system's lifecycle (Mäntymäki et al., 2022).

One positive trend is the rising availability of methodologies for ethics-based auditing and impact assessment. These methods assess AI systems based on a set of ethical standards and can assist organizations in pre-emptively identifying potential issues before deployment. Ethics audits can evaluate the following: societal impact, privacy protection, fairness, transparency and accountability. These can serve as a link between abstract ethical concepts and actual governance procedures (Mokander & Floridi, 2024).

Another key finding is the importance of Explanatory Artificial Intelligence (XAI). Explainability is emerging as a key requirement for trustworthy AI for its ability to help stakeholders make sense of, assess and question algorithmic decisions. Transparent systems help to ensure accountability, compliance, and trust. Increasingly, governance frameworks impose demands on organisations to explain high impact decisions made with AI involving individuals and communities (OECD, 2024).

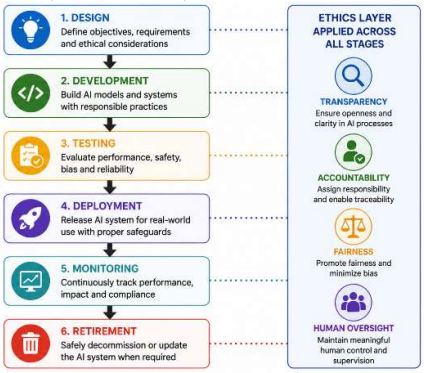

Figure 4. 7 Future AI Governance Lifecycle Model

Figure 4.7 depicts an ethics governance lifecycle framework that integrates ethical oversight into the AI system lifecycle.

The results also underscore the need to pursue inclusive and participatory governance. Technological entities or governmental bodies should not be the sole rulers of ethical AI. Participation of civil society, academics, professional groups, and communities affected is essential for effective governance. Inclusive governance fosters legitimacy, promotes transparency and ensures that different views are taken into account in policy making (UNESCO, 2024).

In addition, international collaboration will become more critical with AI systems crossing borders. The United Nations has put forward the importance of international cooperation, development of standards, capacity building, and governance mechanisms that can address global AI risks. International coordination is especially vital to tackle issues like misinformation, cybersecurity threats, autonomous systems, and cross-border data governance.

Figure 4. 8 Global Ethical AI Governance Ecosystem

Figure 4.8 shows the linkages among governments, international organizations, the private sector, academia, and civil society in a global AI governance ecosystem.

In summary, the results show a need for a shift in focus from reactive to proactive, adaptive, and human-focused regulation of AI in the future. Technological innovation, ethics, the rule of law, and democracy need to be combined in ethical governance. This can help to ensure that the benefits of AI are maximized for society while reducing the risks to human rights, autonomy, fairness and social justice. Responsible AI governance is not just a legal obligation, but also a moral imperative to ensure that technological advancements are grounded in human values and the collective good.

5. CONCLUSIONS, LIMITATIONS AND FUTURE SCOPE

The implications of Artificial Intelligence (AI) in the human decision-making system and philosophical discourse were explored in this study. The study conducted a qualitative analysis of current literature, ethical guidelines, and governance models to recognize the growing role of AI, both technologically and philosophically. This research highlights the socio-technical nature of AI and its potential to shape human actions, institutional decisions, and societal outcomes, moving beyond simply being a computational tool.

The analysis identified that classical theories of ethics, such as utilitarianism, deontological ethics, virtues ethics and ethics of care, remain relevant for the assessment of AI systems. Utilitarian values focus on the efficiency and social good, while deontological values focus on respect for human rights, human dignity and human autonomy. The ethics of caring and virtue ethics also emphasize the moral implications of social welfare and human focus in the process of creating AI systems. These philosophies all show that ethical AI cannot simply be judged by technical performance, but it can be judged by moral and social values as well.

The study also confirmed that AI has a profound impact on human decision-making in various areas such as healthcare, finance, education, the legal field, and governance. While AI technologies can enhance efficiency, predictive insights, and decision-making capabilities, they also pose risks of algorithmic bias, lack of transparency, privacy issues, accountability concerns, and limitations on human autonomy. The results suggest that relying too heavily on AI recommendations can diminish critical thinking and judgment, especially when users believe that they give accurate and flawless advice (Leslie et al., 2024).

Another big point of accountability is drawn. The study identified few arguments to support the notion that AI systems are moral agents. But ethical and legal responsibility still lies with developers, organisations, regulators and users who design, deploy and supervise AI systems. As a result, the key to effective governance is to establish clear accountability mechanisms, involve humans meaningfully in the process, and monitor it constantly during the AI lifecycle.

Lastly, the study suggests that the future of AI governance hinges on the effective combination of ethical principles and regulatory frameworks and organizational practices. Increasingly, a number of international organizations have developed principles and guidelines for the ethics of AI, including the OECD AI Principles and the UNESCO Recommendation on the Ethics of AI, which underscore principles of transparency, fairness, accountability, protection of human rights, and human-centred governance. Ethical governance should thus be considered as more of a prerequisite for developing trustworthy and socially beneficial AI systems than an impediment to innovation (OECD, 2024; UNESCO, 2024).

REFERENCES

1. Batool, A., Zowghi, D., & Bano, M. (2024). *Responsible AI governance: A systematic literature review*. arXiv. https://arxiv.org/abs/2401.10896
2. Brand, J. L. M., & Nannini, L. (2023). *Does explainable AI have moral value?* https://arxiv.org/abs/2311.14687

3. Cheong, B. C., et al. (2024). *Transparency and accountability in AI systems*. *Frontiers in Human Dynamics*. https://www.frontiersin.org/journals/human-dynamics/articles/10.3389/fhumd.2024.1421273/full
4. European Commission. (2024). *AI Act: Regulatory framework for artificial intelligence*. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
5. Fritz, A. (2020). *Moral agency without responsibility? Analysis of three models of agency*. https://de-ethica.com/article/view/3116
6. Google AI. (2025). *How we're making AI helpful for everyone*. https://ai.google
7. Khan, A. A., Badshah, S., Liang, P., Khan, B., Waseem, M., Niazi, M., & Akbar, M. A. (2021). *Ethics of AI: A systematic literature review of principles and challenges*. arXiv. https://arxiv.org/abs/2109.07906
8. Leslie, D., Rincon, C., Briggs, M., Perini, A., Jayadeva, S., Borda, A., Bennett, S. J., Burr, C., Aitken, M., Katell, M., & Fischer, C. (2024). *AI ethics and governance in practice: An introduction*. arXiv. https://arxiv.org/abs/2403.15403
9. Manoli, A., Pauketat, J. V. T., & Anthis, J. R. (2024). *The AI double standard: Humans judge all AIs for the actions of one*. https://arxiv.org/abs/2412.06040
10. Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). *Putting AI ethics into practice: The hourglass model of organizational AI governance*. arXiv. https://arxiv.org/abs/2206.00335
11. Mokander, J., & Floridi, L. (2024). *Operationalising AI governance through ethics-based auditing: An industry case study*. https://arxiv.org/abs/2407.06232
12. OECD. (2024). *AI Principles*. Organisation for Economic Co-operation and Development. https://www.oecd.org/en/topics/sub-issues/ai-principles.html
13. OECD. (2024). *Governing with artificial intelligence*. Organisation for Economic Co-operation and Development. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/governing-with-artificial-intelligence_f0e316f5/26324bc2-en.pdf
14. OECD. (2024). *Recommendation of the Council on Artificial Intelligence*. OECD Legal Instruments. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449
15. OECD. (2024, May 3). *OECD updates AI Principles to stay abreast of rapid technological developments*. https://www.oecd.org/en/about/news/press-releases/2024/05/oecd-updates-ai-principles-to-stay-abreast-of-rapid-technological-developments.html
16. Radanliev, P. (2025). *AI ethics: Integrating transparency, fairness, and privacy*. https://www.tandfonline.com/doi/full/10.1080/08839514.2025.2463722
17. Schmitz, C., & Bryson, J. (2025). *A moral agency framework for legitimate integration of AI in bureaucracies*. https://arxiv.org/abs/2508.08231
18. Sun, N., Miao, Y., Jiang, H., Ding, M., & Zhang, J. (2024). *From principles to practice: A deep dive into AI ethics and regulations*. arXiv. https://arxiv.org/abs/2412.04683
19. Takeshita, M., Rafal, R., & Araki, K. (2023). *Towards theory-based moral AI*. https://arxiv.org/abs/2306.11432
20. The Guardian. (2024, December 6). *Revealed: Bias found in AI system used to detect UK benefits fraud*. https://www.theguardian.com/society/2024/dec/06/revealed-bias-found-in-ai-system-used-to-detect-uk-benefits
21. UNESCO. (2024). *Ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics
22. UNESCO. (2024). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence
23. United Nations Advisory Body. (2024). *Recommendations for governing artificial intelligence*. Reuters. https://www.reuters.com/technology/artificial-intelligence/un-advisory-body-makes-seven-recommendations-governing-ai-2024-09-19/