ETHICAL DILEMMA IN ARTIFICIAL INTELLIGENCE: ANALYZING AI DECISION MAKING FROM A MORAL PERSPECTIVE

BY

MUBASHMAT OCHUWA IBRAHIM (MISS)

ART2101096

DEPARTMENT OF PHILOSOPHY

FACULTY OF ARTS

UNIVERSITY OF BENIN

BENIN CITY

OCTOBER, 2025

ETHICAL DILEMMA IN ARTIFICIAL INTELLIGENCE: ANALYZING AI DECISION MAKING FROM A MORAL PERSPECTIVE

BY

MUBASHMAT OCHUWA IBRAHIM (MISS)

ART2101096

AN ORIGINAL ESSAY SUBMITTED TO THE DEPARTMENT OF PHILOSOPHY, FACULTY OF ARTS, UNIVERSITY OF BENIN, BENIN CITY. IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF BACHELOR OF ART (B.A) DEGREE IN PHILOSOPHY.

OCTOBER, 2025

CERTIFICATION

This is to certify that this project work titled: **ETHICAL DILEMMA IN ARTIFICIAL INTELLIGENCE: ANALYZING AI DECISION MAKING FROM A MORAL PERSPECTIVE** was carried out by **MUBASHMAT OCHUWA IBRAHIM (MISS)** with matriculation number **ART2101096** of the Department of Philosophy, Faculty of Arts, University of Benin, Benin-City.

_____ _____

DR. W. T OSEMWENGIE **DATE**
(Project Supervisor)

_____ _____

DR. W. T OSEMWENGIE **DATE**
(Ag. Head of Department)

DEDICATION

With a grateful heart, I dedicate this work to God Almighty for his unconditional love and provision over my life and also to my lovely parents for their financial support, prayers and love through out my stay in school.

ACKNOWLEDGMENTS

With a grateful heart, I thank the Almighty, God who in His infinite mercies saw me through my years of study by his Divine protection and provisions and has made this work possible.

The success and completion of my project would have not been possible without the professionalism of my highly esteemed Supervisor who doubled as the Head of Department Dr W. T Osemwengie for his unrelenting efforts and patience in reading, scrutinizing and correcting my work.

Also, I acknowledge Dr Prof. Peter Omonzejele, Prof. Gorge Uzoma Ukagba for his advice, patience in the course of my sojourning at the University of Benin, Prof. Anthony Afe- Asekhauno, Prof. Sylvester Odia, Prof. Felix Airoboman, Dr. V. E. Obinyan, and Dr. Christopher Osawaru, Dr. Emanuel Asia, Dr. J. N. Odigie, Dr. Sylvester Apologun, Dr. Paul, Michael, Dr. Victor Jeko, and Mr. Ibrahim Abdulahi. Thank you all for the impact.

I specially declare my highest gratitude and appreciation toward my friends and my roommates who have supported me during my time of working on this project Favor, Melody, Dayo, Isioma, Sonia thank you all for your support.

My sincere appreciation goes to my ever loving and caring Parents Mr/Mrs Ibrahim for their love, support, prayers and best wishes for me which has kept me moving. I also appreciate my wonderful siblings for their supportive hand and warm heartedness towards me. My prayers to you all is that God will richly bless you all in Amen.

I really appreciate my entire coursemates especially Joel (class rep) whose presence, assistance and love has been unending. Though I cannot mention all your names due to space constraint, this research project would not be complete without acknowledging you all and also for the fun we had together during this four year of study. God bless you all.

TABLE OF CONTENTS

CHAPTER TWO: THE IDEA OF ETHICS AND MORALITY

CHAPTER THREE: ETHICAL DILEMMA IN THE USE OF ARTIFICIAL INTELLIGENCE FOR MORAL DECISION MAKING

CHAPTER FOUR: EVALUATION, RECOMMENDATIONS AND CONCLUSION

ABSTRACT

The significance of ethics in Artificial Intelligence (AI) can not be overstated, as it encompasses the foundational principles guiding the responsible creation, deployment and management of AI technologies. As AI systems increasingly permeates every aspect of our lives from healthcare and education to security and entertainment, their decisions and actions have profound implications not only on individual rights and privacy but also on societal norms and values. Ethical considerations in AI are paramount to ensure that these technologies enhance human well-being, uphold fairness rather than perpetuate biases, exacerbate inequalities or undermine democratic institutions. The importance of AI ethics lies in its ability to provide a framework for navigating the complex moral dilemma presented by AI, such as balance between innovation and regulation, the protection of individual privacy versus the benefits of big data and the protection of AI misuse. This project explores AI decision making from an ethical perspective, examining issues such as bias, accountability, transparency and fairness. Through case studies and theoretical analysis, it evaluates how AI systems navigate morally complex situations and the extent to which they align with human ethical principles. This study also discusses existing ethical framework such as Unitarianism, deontology and virtue ethics. Ultimately, the goal is to highlight the need for responsible AI development and governance to ensure that AI driven decisions uphold ethical standards and benefits society as a whole.

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND TO THE STUDY

Artificial Intelligence (AI) is a branch of computer science that focuses on creating machines capable of performing tasks that typically requires human intelligence. These tasks include reasoning, problem solving, learning, perception, language understanding and decision making. According to Oxford Learner's Dictionary, Artificial Intelligence is the study and development of computer systems that can copy intelligent human behavior.[1]

The history of artificial intelligence (AI) development is a fascinating journey that spans several decades, tracing back to the mid-20th century when the concept of creating intelligent machines first captured the imagination of scientists and philosophers. The formal inception of AI as a scientific discipline is often attributed to the 1956 Dartmouth Conference, where pioneers like John McCarthy, Marvin Minsky, Allen Newell, and Herbert A. Simon set the ambitious goal to explore how machines could be made to simulate aspects of human intelligence. This period saw the development of early AI

---

[1] Horny, A. S. (2005), *Oxford Learner's Dictionary*, 8th Edition, (Oxford: Oxford University Press), p 101.

programs, such as the Logic Theorist and ELIZA, which demonstrated problem-solving and natural language processing capabilities.[2]

According to John MC Cathy, who is considered one of the founding figures of Artificial intelligence, MC Cathy defined AI as the "The science and engineering of making intelligent machines, especially intelligent computer programs, his definition emphasizes AI as both a scientific discipline and an engineering challenge focused on replicating human like intelligence.[3]

For Akande, Michael, Artificial intelligence was introduced by Norbert Winner around 1948. This concept was used to describe at that period a new wave of seeing machine as human counterparts. Scholars assumed that if human possess natural intelligence then these machine counterparts must possess artificial intelligence. They are considered artificial because computers or machines are not natural or biological creatures. In this light, artificial intelligence is the science of digital machine doing the sorts of things that are done by human minds.[4] The current state of artificial intelligence (AI) technologies is characterized by rapid advancements and widespread integration into various sectors of

---

[2] Liao S. M., eds, (2020), *Ethics of Artificial Intelligence*, (Oxford: Oxford University Press), p. 13.

[3] McCarthy, J. (2007). *What is Artificial Intelligence?*. Retrieved from http://jmc.stanford.edu/articles/whatisai.html

[4] Akande, M.A. (2011), Artificial Intelligence and the limits of Epistemic Justification *LASU Journal of Philosophy*.

society, leading to significant societal impacts. AI systems, powered by machine learning algorithms and vast amounts of data, are now capable of performing complex tasks with precision and efficiency that rival or surpass human capabilities in some areas. These technologies have found applications in healthcare, where they assist in diagnosing diseases and personalizing treatment plans; in finance, through algorithmic trading and fraud detection; in transportation, via autonomous vehicles and smart traffic management systems; and in everyday consumer products, including virtual assistants and recommendation algorithms.

There are a number of different forms of learning as applied to artificial intelligence. The simplest is learning by trial and error. For example, a simple computer program for solving mate-in-one chess problems might try moves at random until mate is found. The program might then store the solution with the position so that, the next time the computer encountered the same position, it would recall the solution. This simple memorizing of individual items and procedures, known as rote learning, is relatively easy to implement on a computer.

More challenging is the problem of implementing what is called generalization. Generalization involves applying past experience to analogous new situations. For example, a program that learns the past tense of regular English verbs by rote will not be able to produce the past tense of a word such as jump unless the program was previously presented with jumped, whereas a program that is able to generalize can learn the “add -

ed" rule for regular verbs ending in a consonant and so form the past tense of jump on the basis of experience with similar verbs.

Some authors offer the Turing test as a definition of intelligence. However, the mathematician and logician Alan Turing himself pointed out that a computer that ought to be described as intelligent might nevertheless fail his test if it were incapable of successfully imitating a human being. For example, ChatGPT often invokes its status as a large language model and thus would be unlikely to pass the Turing test. If an intelligent entity can fail the test, then the test cannot function as a definition of intelligence. It is even questionable whether passing the test would actually show that a computer is intelligent, as the information theorist Claude Shannon and the AI pioneer John McCarthy pointed out in 1956. Shannon and McCarthy argued that, in principle, it is possible to design a machine containing a complete set of canned responses to all the questions that an interrogator could possibly ask during the fixed time span of the test. Like PARRY, this machine would produce answers to the interviewer's questions by looking up appropriate responses in a giant table. This objection seems to show that, in principle, a system with no intelligence at all could pass the Turing test.[5]

---

[5]Copeland, J. (2000), "What is Artificial Intelligence?" *Alan Turing.Net*, © Copyright B.J. Copeland.

1.2 STATEMENT OF THE PROBLEM

Despite the substantial benefits, the proliferation of AI technologies also raises critical societal concerns. The potential for job displacement due to automation, issues of privacy and surveillance arising from data-centric AI applications, and the amplification of biases in decision-making algorithms highlight the dual-edged nature of AI's impact on society. Furthermore, the increasing reliance on AI systems underscores the importance of addressing ethical considerations, such as transparency, accountability, and fairness, to ensure these technologies contribute positively to societal well-being. As AI continues to evolve, its societal impact will likely deepen, necessitating ongoing dialogue, policy development, and ethical considerations to harness its potential while mitigating adverse effects.

AI systems are now responsible for diagnosing diseases, approving loans, predicting criminal activity, and even making life-or-death decisions in autonomous vehicles. While AI is celebrated for its ability to enhance efficiency, objectivity, and scalability, its increasing autonomy raises profound ethical concerns about fairness, accountability, and the moral implications of delegating critical decisions to machines. One of the most pressing issues in AI ethics is algorithmic bias. AI systems learn from historical data, which often reflects societal inequalities. As a result, AI-driven hiring tools have shown gender and racial biases, while predictive policing algorithms have disproportionately targeted minority communities. These biases challenge the assumption that AI decisions

are objective and impartial, raising concerns about fairness and social justice. If AI systems are unintentionally reinforcing systemic discrimination, how can they be ethically designed and regulated?[6]

Another major concern is the lack of transparency and explainability in AI decision-making. Many AI models, particularly deep learning systems, operate as "black boxes," where even their creators struggle to explain how decisions are made. This opacity makes it difficult to ensure accountability—who should be held responsible when an AI system makes a harmful decision? Traditional legal and ethical frameworks were built around human decision-making, but they struggle to assign moral or legal responsibility when decisions are made by an AI.

Additionally, AI-driven automation raises questions about autonomy, control, and the role of human oversight. In some domains, AI systems are designed to work alongside humans, providing recommendations or assisting with decision-making. However, in other cases, AI is given full autonomy, as seen in self-driving cars and autonomous weapons. This shift raises serious moral dilemmas: Should an autonomous vehicle prioritize the life of its passengers over pedestrians in an unavoidable accident? Should AI-powered military drones have the authority to make lethal decisions without human intervention? These ethical challenges highlight the difficulty of embedding human values into AI systems.

---

[6] Stanford Encyclopedia of Philosophy. (2020). *Ethics of Artificial Intelligence and Robotics.*

From a broader philosophical perspective, AI challenges traditional ethical theories that have guided human decision-making for centuries. Utilitarianism, which emphasizes maximizing overall benefit, may justify AI decisions that sacrifice individual rights for the greater good. Deontological ethics, which focuses on following moral rules and duties, may struggle when AI systems must make trade-offs in complex moral situations. Virtue ethics, which emphasizes moral character, is particularly difficult to apply to AI, as machines lack emotions, intentions, and moral consciousness. These dilemmas raise the question: Can AI ever truly align with human ethical principles, or will it always operate in a morally ambiguous space?

Given the growing influence of AI in decision-making, addressing these ethical concerns is crucial. Without careful consideration of AI's moral implications, society risks deploying technologies that inadvertently cause harm, reinforce discrimination, or operate without clear accountability. This study seeks to explore these challenges through the lens of moral philosophy, legal responsibility, and technological ethics, aiming to provide insights into how AI can be developed and regulated to ensure ethical decision-making. Hence there is a need to consider the ethical implication of Artificial Intelligence in this study.

1.3 PURPOSE OF THE STUDY

The purpose of the study of the ethical dilemma in Artificial Intelligence are:

1. Ensures that Artificial Intelligence systems are developed and used responsibly, ethically in a way that aligns with human values and societal well-being.

2. Examines how Artificial Intelligence algorithm make decisions including their reliance on data, machine learning models and predefined ethical frameworks.

3. Identity and analyze ethical dilemmas Artificial Intelligence systems face such as bias, accountability and privacy concerns.

This study aims to build the gap between AI development and ethical considerations ensuring responsibile Artificial Intelligence implementation for the benefit of humans and society.

1.4 SIGNIFICANCE OF THE STUDY

The significance of this study lies in its potential to address critical concerns surrounding AI ethics and responsible implementation.

The significance of this study are:

1. Help researchers understand moral challenges in Artificial Intelligence decision making.

2. Help in designing Artificial Intelligence systems that align with human values and moral reasoning.

3. Connects ethical theories (utilitarianism, deontology) with Artificial Intelligence decision making.

4. Highlights the importance of Artificial Intelligence to improve trust and fairness.

5. Ensure Artificial Intelligence benefits society while minimizing harm and discrimination.

6. Encourage the integration of ethical principles into Artificial Intelligence design.

1.5 SCOPE AND LIMITATIONS OF THE STUDY

The study focuses on the ethical dilemmas associated with Artificial Intelligence decision making from a moral perspective. It examines how AI systems navigate ethical challenges including bias, fairness, accountability, transparency and privacy. This research explores various philosophical frameworks such as utilitarianism, deontology and virtue ethics and their applicability in Artificial Intelligence ethics.

Additionally, the study analyzes real world case studies in areas such as healthcare, law enforcement, finance and autonomous system to illustrate the moral complexities in AI driven decision.

The study also evaluates existing ethical guidelines and policies designed to regulate AI and propose strategies for ensuring responsible AI development. By addressing these

ethical concerns, the research aims to contribute to the ongoing debate on AI governance and the need for human oversight in AI decision making.

This study is primarily theoretical and analytic, relying on literature review and case studies rather than experimental AI development or testing.

The study is limited by the fact that ethical principles and moral perspectives can be subjective and vary across cultures and contexts.

Many AI algorithms used in industries are proprietary and lack transparency making it difficult to fully access their decision making processes.

Ethical AI standards differ across countries and this study may not comprehensively cover all global perspective and regulation.

Artificial Intelligence technology is advancing quickly meaning that new ethical challenges may emerge that are not covered in this study. Despite these limitations, this research provides valuable insights into the ethical implications of AI decision making and contributes to the broader discussion on responsible AI development.

1.6 METHODOLOGY

Methodology refers to the systematic approach, strategies and techniques used in a research study to collect, analyze and interpret data. It outlines the procedures that guide the research process ensuring that the study is conducted in a structured and reliable manner. This project adopts the method of critical analysis which has to do with the

breaking down of concepts into smaller units. This method will enable us to examine the concepts of ethics, Artificial Intelligence and decision making.

1.7 DEFINITION OF TERMS

The pursuit of clarity and precision demands that we establish clear definitions of key terms to be utilized throughout this study.

Moral: The term moral is concerned with the principles of right and wrong behavior and the goodness or badness of human character.[7] In philosophical usage, "moral" pertains to standards of behavior or beliefs regarding what is and is not acceptable for individuals or society.

Artificial Intelligence: According to Yann LeCunn, Chief AI Scientist at Meta, Artificial Intelligence is “any information technology capable of solving complex problems that would normally be attributed to humans and animals”.[8]

---

[7] Hornby, A. S. (2005), *Oxford Advanced Learner's Dictionary 8th Edition*, (Oxford: Oxford University Press), p. 90.

[8] Le Cun, Yann. Destination AI: Introduction to Artificial Intelligence." OpenClassrooms, openclassrooms.com/en/courses/7078811-destination-ai-introduction-to-artificial-intelligence. Accessed 13-06-2025.

Freewill: According to the Oxford English Dictionary, freewill (or free will) is defined as the power of acting without the constraint of necessity or fate; the ability to act at one's own discretion. In philosophy, free will refers to the capacity of rational agents to choose a course of action among various alternatives. It implies that individuals are responsible for their actions because they have control over their decisions.[9]

Epistemological: According to the Oxford English Dictionary, the term "epistemological" is defined as relating to the theory of knowledge, especially with regard to its methods, validity, and scope. It stems from the root word epistemology, which is the branch of philosophy that investigates the nature, origin, and limits of human knowledge.[10]

1.8 LITERATURE REVIEW

As artificial intelligence (AI) systems increasingly make decisions with significant social, economic, and even life-or-death consequences, questions about the ethical dimensions of AI decision-making have become urgent. Researchers, ethicists, technologists, and policymakers are grappling with how AI can and should behave in ways that align with human moral values. This literature review critically examines the major scholarly contributions to the understanding of ethical dilemmas in AI, with a focus on how AI decision-making is analyzed from a moral perspective.

---

[9] Hornby, A. S. *Op. Cit.*, p. 129.
[10] Hornby, A. S. *Op. Cit.*, p. 100.

Several scholars have proposed the application of traditional ethical theories to AI systems.

Brynjolfsson, Erik & Andrew McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*[11], offers a broad, empirically informed account of how digital technologies and automation reshape economies, labor markets, and social institutions. Their central claim, that advances in computing create discontinuities in productivity and social organization, frames why ethical analysis is urgent: technological change alters the contexts in which moral decisions and responsibilities arise. For studies on AI and moral agency, the book supplies socio-economic background explaining how automated decision systems come to play outsized roles in workplaces, public policy, and everyday life. It also raises normative questions about distribution, justice, and the responsibilities of designers, firms, and states to mitigate harms. While not a work of normative theory, its interdisciplinary approach helps bridge philosophical concerns about autonomy and responsibility with the material realities that make those concerns pressing. In short, it is essential reading for situating philosophical and ethical debates about AI within contemporary socio-economic transformations.

[11] Brynjolfsson, E. & McAfee, A. (2014), *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, (New York: W.W. Norton), p. 34.

M. Coeckelbergh *AI Ethics*[12] provides a concise, philosophically rich treatment of ethical issues raised by AI, moving beyond narrow technical fixes to probe questions of trust, responsibility, dignity, and political economy. He draws on a range of ethical traditions (virtue ethics, deontology, consequentialism) and connects them to concrete design and governance problems, emphasizing relational and social-ontological perspectives. His work is particularly useful for scholars seeking conceptual tools to analyze human–AI interaction: he reframes ethical problems as questions about how technologies reconfigure social practices and moral perceptions. Coeckelbergh's emphasis on social embeddedness offers a corrective to purely individualistic accounts of moral agency. The book's pragmatic orientation, linking ethical reflection to policy and design, makes it valuable for translating philosophical diagnosis into institutional recommendations.

J, Copeland's "What is Artificial Intelligence?"[13] online essay is a clear, historically informed introduction to the conceptual foundations of AI, surveying Turing's work, symbolic and statistical paradigms, and debates over machine cognition. It helps researchers frame core definitional questions (what counts as AI, what counts as intelligence) that underlie ethical debates: misunderstandings about AI's capacities and limits often produce misplaced moral expectations or anxieties. While exploring moral decision-making by machines, Copeland's taxonomy clarifies which kinds of systems

---

[12] Coeckelbergh, M. (2020), *AI Ethics*, (London: MIT Press), p. 19.

[13] Copeland, J. (2000), "What is Artificial Intelligence?" *Alan Turing.Net*, © Copyright B.J. Copeland.

(rule-based, learning-based, hybrid) are likely to raise which ethical issues. His historical perspective also assists normative inquiry by reminding scholars that technological categories change over time, an important caution when making claims about agency, responsibility, or the moral status of machines. Although not a normative text, it is indispensable background for any rigorous discussion of AI ethics.

W. K. Frankena, *Ethics*[14], presents classic moral theories (utilitarianism, Kantian deontology, virtue ethics) with clarity and philosophical rigor, making it a reliable foundation for comparative ethical analysis. His systematic exposition of Kantian ethics, its premises, structure, and key objections, provides the conceptual machinery needed to examine questions about autonomy, duty, and moral law. For a study comparing Kant's responses to determinism with modern debates sparked by AI, Frankena's clear delineation of deontological commitments is invaluable. Moreover, his discussion of criteria for moral justification and moral theory evaluation offers methodological guidance for assessing contemporary proposals like value alignment or algorithmic consequentialism. Frankena thus serves both as a primer on established moral frameworks and as a critical tool for interrogating how those frameworks apply (or fail to apply) to technological contexts.

[14] Frankena, W. K. (1973). *Ethics* (2nd ed.). (Englewood Cliffs, NJ: Prentice-Hall), p. 34.

I. Gabriel's, *Artificial Intelligence, Value and Alignment*[15] addresses the technical and philosophical problem of value alignment: how to ensure that AI systems act in ways compatible with human values. He critically examines different strategies (pre-specified objectives, learning from human behavior, normative constraints) and highlights conceptual pitfalls such as value pluralism, specification gaming, and the opacity of learned models. The paper is especially pertinent to projects that interrogate whether algorithmic systems can be genuine moral agents or should instead be constrained to support human moral agency. Gabriel's balanced engagement with both philosophical theory and machine-learning practice makes his article a key bridge between abstract ethical concerns (about what counts as a value) and pragmatic engineering solutions. It provides sharp analytical tools for evaluating claims that alignment is merely a technical challenge rather than a deeply normative problem.

S. Gbadegesin, *African Philosophy: Traditional Yoruba Philosophy and Contemporary African Realities*[16]**,** presents Yoruba philosophical categories (notably ideas about personhood, community, and moral responsibility) and explores how traditional thought informs contemporary ethical practice in Africa. His work is crucial for situating debates about autonomy, moral agency, and accountability within African normative frameworks

---

[15] Gabriel, I. (2020), "Artificial Intelligence, Value and Alignment." *Minds and Machines*, vol. 30.

[16] Gbadegesin, S. (1991). *African Philosophy: Traditional Yoruba Philosophy and Contemporary African Realities*, (Chicago: Gateway Press), p. 54.

rather than importing exclusively Western paradigms. For research on AI and morality in African contexts, Gbadegesin's analyses prompt critical questions: how do communalist conceptions of personhood change our understanding of responsibility for technological harms, and what do indigenous notions of wisdom and leadership imply about the role of traditional institutions in governing AI? Incorporating Gbadegesin helps ensure that ethical appraisals of AI are culturally sensitive and attentive to plural moral vocabularies.

K. Gyekye, *Tradition and Modernity: Philosophical Reflections on the African Experience*[17], Gyekye offers a measured account of how African societies negotiate the tension between traditional values and modern institutions, arguing that tradition can be a source of normative resources rather than merely an obstacle to development. His work contributes to debates about the compatibility of local moral frameworks with global ethical norms, which is highly relevant when considering transnational deployment of AI systems in Africa. For scholars examining Kantian notions of autonomy alongside African communitarian ethics, Gyekye's nuanced stance furnishes conceptual space for dialogue: rather than subsuming African thought under Western categories, he points to hybrid, context-sensitive syntheses. This makes his book useful for policy-oriented recommendations that seek culturally legitimate governance models for AI.

---

[17] Gyekye, K. (1997). *Tradition and Modernity: Philosophical Reflections on the African Experience*, (Oxford: Oxford University Press), p. 46.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, *Ethically Aligned Design*[18]. The IEEE report aggregates expert guidance and operational principles for the ethical design and governance of autonomous systems, emphasizing human well-being, transparency, accountability, and governance structures. Though institutional and policy-focused rather than purely philosophical, *Ethically Aligned Design* is an influential practical resource: it translates philosophical tenets (e.g., respect for persons, fairness) into design requirements, audit practices, and governance proposals. For projects bridging Kantian moral concepts and AI practice, IEEE's work is useful in showing how abstract duties and rights might be operationalized in engineering and institutional contexts. It also highlights practical constraints and trade-offs, for example, between explainability and performance, that philosophical arguments must account for when they inform design.

R. James, *The Elements of Moral Philosophy*[19] is a compact, pedagogical introduction to major ethical theories and contemporary moral issues, useful for grounding readers new to moral philosophy. Its accessible discussions of dilemmas, applied ethics, and contrasting normative frameworks help structure comparative analyses (for example,

---

[18] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. IEEE.

[19] James, R. (2012), *The Elements of Moral Philosophy*, 7th ed., (New York: Mc Graw-Hill Education), p. 17.

Kantian versus utilitarian responses to automation and machine decision-making). James's work provides clear summaries, illustrative examples, and critical questions that can guide deeper engagement with primary philosophical sources. It is therefore a convenient reference point to orient readers before introducing more specialized literature on AI and tradition.

I. Kant, *Groundwork of the Metaphysics of Morals*,[20] is the central canonical text for understanding autonomy, moral law, and the categorical imperative, concepts that animate debates about moral agency, responsibility, and dignity in the age of AI. His insistence on rational autonomy and duty provides a stringent standard against which to evaluate proposals that delegate moral judgment to machines. The *Groundwork* also frames the determinism debate through Kant's distinction between the phenomenal and noumenal, a distinction many contemporary scholars appeal to when arguing for a space for moral responsibility in causally structured worlds. For any rigorous study of AI and moral decision-making, especially one that seeks to interrogate human freedom, accountability, or the normative limits of algorithms.

[20] Kant, I. (1997), *Groundwork of the Metaphysics Morals*, trans. Mary Gregor.(Cambridge:
Cambridge University Press), p. 91.

CHAPTER TWO

THE IDEA OF ETHICS AND MORALITY

2.1 WHAT IS ETHICS AND MORALITY?

Ethics is a core area of philosophy that deals with principles guiding human behavior and judgment. It provides a rational foundation for determining what is right or wrong, just or unjust, virtuous or immoral. Unlike mere social customs or legal rules, ethics is concerned with how human beings ought to act based on reason, reflection, and shared values. Ethics plays a crucial role in shaping human conduct, regulating our actions, and guiding interpersonal and social relationships. It functions as a critical lens through which human beings evaluate what ought to be done, not only in isolated personal decisions but also in complex societal arrangements. Within philosophy, ethics offers a space for rational inquiry into values and norms, helping societies to articulate visions of justice, goodness, and the ideal human life. It is neither a rigid set of instructions nor merely a cultural inheritance; it is a dynamic discipline that invites continuous reflection on the nature of right and wrong in an ever-changing world. The term ethics comes from the

Greek word ethikos, meaning “moral” or “relating to character,” and is derived from ethos, meaning “custom” or “habit.” In philosophical usage, ethics refers to the study of morality, that is, the rational examination of right and wrong, good and bad conduct. Ethics is not simply a theoretical subject; it influences personal choices, social interactions, and public life. Throughout history, philosophers have treated ethics as a serious inquiry into the conditions for a good and meaningful life. It remains essential for fostering moral clarity, civic responsibility, and individual integrity. Its subject consists of the fundamental issues of practical decision making, and its major concerns include the nature of ultimate value and the standards by which human actions can be judged right or wrong. This capacity for self-reflection and judgment is what distinguishes moral agents from mere automatons. As Immanuel Kant argues in *Groundwork of the Metaphysics of Morals* (1785), moral agency rests on the ability to act according to maxims that one can will as universal laws. He insists that human beings must always be treated as ends in themselves, not merely as means to an end.[21] This notion lays the foundation for respect for human dignity, a core concept in contemporary moral philosophy. In contrast to Kant’s deontological ethics, Aristotle offers a virtue-based perspective, one grounded in the development of moral character. In his *Nicomachean Ethics*, Aristotle posits that ethical living is about cultivating virtues such as courage, temperance, justice, and prudence through practice and habituation. According to him, the goal of human life is

---

[21] Kant, I. (1997), *Groundwork of the Metaphysics Morals*, trans. Mary Gregor, (Cambridge:
Cambridge University Press), pp. 36-49.

eudaimonia a state of flourishing or well-being that arises from living in accordance with reason and virtue.[22] Virtue ethics moves away from abstract rules and focuses on the person's moral disposition, making it particularly relevant in contexts that value moral integrity and communal harmony. Moral objectivists like John Rawls defend a system of justice based on fairness and equality. In *A Theory of Justice*, Rawls (1971) proposes the "original position" a hypothetical situation where individuals, behind a "veil of ignorance," agree on the basic structure of a just society. The goal is to ensure that social and political institutions do not privilege some groups over others and that all persons are treated with equal moral worth.[23] This framework has been influential in debates about ethics, rights, and development across both Western and African societies.

Furthermore, Oladipo contends that traditional African moral values can be harnessed to address modern social problems, provided they are critically reassessed and adapted. He warns against both wholesale rejection and blind acceptance of tradition. Instead, he advocates for a dynamic ethical consciousness that is faithful to its roots but responsive to contemporary realities.[24]

---

[22] Aristotle, (1999), *Nicomachean Ethics*, trans. Terence Irwin, (London: Hackett Publishing), pp. 28-35

[23] Rawls, J. (1971). *A Theory of Justice*, (London: Harvard University Press), p. 45.

[24] Oladipo, O. (2006). *Modernity and African Philosophy*, (London: Hope Publications), pp. 92-93.

Ethics also plays a central role in professional contexts from medicine and education to law and business. In medical ethics, for instance, principles such as autonomy, beneficence, non-maleficence, and justice guide healthcare providers in making difficult decisions. In legal ethics, fairness, impartiality, and integrity are paramount. Professional codes of conduct are not just bureaucratic requirements but ethical instruments designed to uphold public trust and ensure responsible practice. Ethics is not only concerned with right action but also with moral formation and development. It raises the question: How can individuals become morally upright persons? This is where moral education becomes essential. Morality is one of the central concepts in ethics and philosophy. It refers to the principles, norms, and standards that govern human behavior concerning what is considered right or wrong, good or bad. Morality guides actions, intentions, and social behavior within cultures and societies. Morality comes from the Latin word, Moralis which means conduct or custom or way of life .According to Immanuel Kant in Groundwork for the Metaphysics of Morals, morality is not based on outcomes or consequences but on duty. He asserts "Morality is not the doctrine of how we may make ourselves happy, but how we may make ourselves worthy of happiness."[25] This definition positions morality not merely as a social tool but as an intrinsic obligation grounded in rationality and moral duty. Moral understanding grows with maturity.

Lawrence Kohlberg, in his theory of moral development, describes stages of moral growth:

---

[25] Kant, I. *Op. Cit.,* p. 62.

1.Pre-conventional: Obedience based on punishment and reward.

2.Conventional: Upholding laws and social order.

3.Post-conventional: Guided by internal moral principles and universal ethics.

This theory shows that morality is a developmental process influenced by cognitive and social growth.

James Rachels in The Elements of Moral Philosophy explains that morality is:

"...at the very least, the effort to guide one's conduct by reason that is, to do what there are the best reasons for doing, while giving equal weight to the interests of each individual affected by one's decision."[26]

In today's rapidly changing world, the concept of morality faces complex and sometimes unprecedented challenges. While traditional moral systems were built around religion, culture, or philosophical reasoning, modern society introduces new dilemmas that require rethinking, adaptation, and sometimes the creation of entirely new ethical frameworks.

1. Technological Advancements and Moral Dilemmas

Modern technology especially Artificial Intelligence (AI), biotechnology, and the internet has raised questions that earlier moral systems never had to consider.

[26] Rachels, J. (2012). *The Elements of Moral Philosophy*, 7th ed., (New York: MC Graw-Hill Education), p. 11.

Artificial Intelligence: Machines can now make decisions that affect lives (e.g., self-driving cars, AI in warfare, or AI in healthcare). The moral question is: Can machines have moral responsibility?

For example, if an autonomous car kills a pedestrian to save its passengers, who is morally accountable the engineer, the AI, or the manufacturer?

Biotechnology and Genetics: With the rise of gene editing (e.g., CRISPR), we can now manipulate human DNA. But should we? Is it moral to eliminate disabilities through genetic engineering? What about "designer babies"?

These questions fall into what ethicist Peter Singer calls the realm of “new ethics,” where:

"We are dealing with acts never contemplated by traditional ethical systems."[27]

2. Globalization and Cultural Relativism

As the world becomes more interconnected through migration, trade, and digital communication, we encounter diverse moral perspectives.

Cultural Relativism: Some argue that morality is culturally dependent that what is moral in one society may be immoral in another (e.g., views on polygamy, LGBTQ+ rights, or freedom of speech).

---

[27] Singer, P. (2011). *Practical Ethics*, 3rd ed., (Cambridge: Cambridge University Press), p. 2.

Moral Universalism: Others advocate for global human rights standards, arguing that certain moral truths (e.g., against torture, slavery) should apply to all humans regardless of culture. The tension between these positions forces modern morality to balance respect for diversity with the defense of universal values.

3. Economic and Social Justice

Modern morality also concerns itself with equality, justice, and rights, particularly in areas such as:

Income Inequality: Is it moral for CEOs to earn 300 times more than the average worker?

Access to Healthcare and Education: Should these be considered moral rights or privileges?

Thinkers like John Rawls in A Theory of Justice propose that a moral society is one where:

"Social and economic inequalities are to be arranged so that they are... to the greatest benefit of the least advantaged."[28]

Morality in the modern world is dynamic, challenging, and deeply contextual. As technology, global interaction, and social awareness evolve, so too must our moral thinking. A rigid, one-size-fits-all morality may no longer be enough. Instead, modern

[28] Rawls, J. *Op Cit.,* p. 266.

morality demands critical thinking, empathy, and the courage to confront new dilemmas with wisdom and humanity.

2.2 ETHICS AND MORALITY: ANY DIFFERENCE?

Untangling the concept of ethics from that of morality has long posed a thorny philosophical challenge, in philosophical discourse, ethics and morality are often mentioned together, yet scholars have long observed that the two operate at different levels of human thought and practice. While they both concern human conduct and the distinction between right and wrong, their approaches, sources, and purposes diverge in subtle but significant ways.

Morality is closely tied to the lived experiences, traditions, and customs of specific societies. It reflects what people in a particular community actually uphold as acceptable behavior. These norms are often shaped by historical circumstances, religious beliefs, and cultural expectations, making them context-bound. MacIntyre explains that morality is "embedded within traditions" and functions as a socially inherited guide to conduct, passed down and reinforced through communal life.[29] This embedded nature means that what is considered morally acceptable in one culture may be rejected in another. For instance, a practice deeply rooted in the heritage of one society could be deemed objectionable when viewed from the standpoint of another community's moral

[29] MacIntyre, A. (2007). *After Virtue* (3rd ed.), (London: University of Notre Dame Press), p. 38.

framework. Ethics, in contrast, occupies a more reflective and critical space. Rather than simply inheriting or conforming to given norms, ethics involves stepping back to examine those norms, questioning their coherence, and testing their validity against broader or even universal principles. It is the reasoned and systematic evaluation of human conduct, seeking to justify or critique the moral codes that communities follow. Beauchamp and Childress emphasize that while morality might guide what people do within a local framework, ethics attempts to establish standards such as justice, beneficence, and respect for autonomy that transcend local traditions and can be applied universally.[30]

This distinction is also evident in their practical application, morality operates at the level of immediate action it is the lived code that governs daily interactions and decisions. Ethics, however, is more theoretical in orientation. It seeks to understand why certain actions are right or wrong and to determine whether the moral beliefs people hold are logically defensible. Bernard Williams observes that morality can exist without a formal ethical theory[31], meaning that communities can function with shared moral codes even if no one has systematically reflected on them. Ethics, however, cannot function without deliberate thought; it is inherently tied to rational analysis, debate, and argumentation. Another point of divergence lies in their binding authority. Moral norms often draw their

---

[30] Beauchamp, T. L., & Childress, J. F. (2019). *Principles of Biomedical Ethics* (8th ed.), (Oxford: Oxford University Press), p. 3.

[31] Williams, B. (1985). *Ethics and the Limits of Philosophy*, (London: Harvard University Press), p. 6.

force from social pressure, cultural tradition, or religious injunctions. People obey these norms because they are part of the shared fabric of their community, and violating them may lead to social sanction or personal guilt. Ethics, by contrast, derives its authority from reasoned justification. As Rachels and Rachels note, ethical principles appeal to "standards of reason that apply to everyone," and they invite individuals to critically examine whether their actions can be defended through rational argument, rather than merely because "this is how we do things here."[32]

Thus, while morality and ethics are deeply interconnected indeed, ethics often builds upon moral experience they are not identical. Morality reflects the particular, the contextual, and the socially embedded, while ethics seeks the universal, the critically examined, and the logically justified. Together, they form the complex terrain through which human beings navigate questions of right and wrong, with morality grounding human conduct in lived reality, and ethics challenging it through reasoned reflection.

2.3 WHAT IS A MORAL DECISION?

In philosophy, a moral decision is understood as the deliberate act of choosing a course of action that accords with established moral principles, ethical duties, or virtuous ideals, based on rational reflection and moral reasoning rather than impulse or mere personal preference. It involves the consideration of what one ought to do in a given situation,

[32] Rachels, J., & Rachels, S. (2019). *The Elements of Moral Philosophy* (9th ed.). (New York: McGraw-Hill Education), p. 4.

guided by standards of right and wrong, good and bad, justice and injustice, as well as the virtues expected of a morally responsible agent. Moral decisions are normative rather than descriptive they concern what should be done, not merely what is done and they require an evaluative process in which the decision-maker weighs possible actions against principles that can be defended as reasonable and fair.

From the perspective of duty-based ethics, notably represented by Immanuel Kant, the making of a moral decision is governed by adherence to universal moral laws. Kant's categorical imperative demands that one act only on a maxim that one could will to become a universal law, which means that a moral decision is right not because of its outcome, but because it fulfills a duty that applies equally to all rational beings.[33]This approach sees actions such as lying as always wrong, even if the immediate consequences seem beneficial, because lying violates a fundamental duty to truthfulness. In contrast, the utilitarian view, most famously expressed by John Stuart Mill, evaluates moral decisions in terms of their outcomes. For Mill, the morally right decision is the one that promotes the greatest happiness for the greatest number, meaning that the morality of an act depends on its tendency to increase overall well-being and reduce suffering .[34]

Under this view, a decision that breaks a conventional rule might still be moral if it yields better consequences for the majority. Another influential perspective, virtue ethics, rooted

---

[33] Kant. I. *Op Cit.,* p. 52-55.
[34] Mill, J. S. (1998). *Utilitarianism* (R. Crisp, Ed.), (Oxford: Oxford University Press), pp. 14-16.

in the works of Aristotle, interprets a moral decision as one that springs from and cultivates good character. Aristotle emphasizes that moral virtue is acquired through habituation, and that a virtuous person naturally makes the right decisions because they have developed the dispositions such as courage, temperance, and justice that enable them to act in ways that realize human flourishing, or eudaimonia[35]. Thus, moral decision-making here is not simply about following rules or calculating outcomes, but about becoming the kind of person who habitually chooses well.

Within African philosophy, including Nigerian traditions, moral decisions are often shaped by communitarian values that place a strong emphasis on social harmony, mutual responsibility, and the interconnectedness of all members of the community. Kwame Gyekye observes that African moral thought is deeply relational, with decisions evaluated according to how they promote unity, cooperation, and the common good.[36] Segun Gbadegesin, in his analysis of Yoruba ethics, points out that moral decision-making is tied to the cultivation of omoluabi, a concept denoting a person of good character whose actions reflect respect for others, honesty, and a commitment to communal welfare.[37]

---

[35] Aristotle, *Op Cit*., pp. 25-27.
[36] Gyekye, K. (1997). *Tradition and Modernity: Philosophical Reflections on the African Experience*, (Oxford: Oxford University Press), pp. 35-37.

[37] Gbadegesin, S. (1991). *African Philosophy: Traditional Yoruba Philosophy and Contemporary African Realities*. Chicago: Gateway Press), pp. 66-68.

In this framework, a moral decision is not merely an individual calculation but a choice made with an awareness of its impact on the broader social fabric. Moral decision-making also operates in the context of moral dilemmas situations where competing moral principles pull in different directions, making it impossible to satisfy all ethical demands at once.. These dilemmas often require balancing values, such as honesty versus compassion, individual rights versus collective welfare, or short-term harm versus long-term benefit.

In such situations, philosophical reflection helps clarify priorities, uncover hidden assumptions, and ensure that the decision is made with integrity and intellectual honesty. The process may also involve what contemporary ethicists call reflective equilibrium, a method in which specific judgments and general principles are adjusted in light of each other until coherence is reached.

In modern contexts, moral decisions are increasingly tested in areas like biomedical ethics, artificial intelligence, environmental policy, and global justice. The complexity of these issues demonstrates that moral decision-making is not static; it evolves alongside societal changes and technological advancements. Yet, at its core, it remains rooted in the human capacity to reason about what is right, to act deliberately in accordance with that reasoning, and to accept responsibility for the consequences.

Regardless of the philosophical tradition, a common thread is that moral decisions require moral reasoning the capacity to deliberate, to take into account relevant facts and values,

to weigh competing considerations, and to act with a sense of accountability. As William Frankena explains, without such reasoning an act may be morally accidental rather than a genuine moral choice, since true moral decision-making presupposes the agent's awareness of the moral dimensions of the situation and their willingness to be held responsible for the outcome.[38] A moral decision, therefore, is not simply a matter of personal preference or social convention; it is a reasoned judgment about what one ought to do, grounded in ethical principles, shaped by philosophical reflection, and accountable to standards that can be justified both to oneself and to others.

2.4 ETHICAL THEORY: CLASSICAL AND MODERN ETHICAL THEORY

Ethical theories provide systematic frameworks for evaluating human conduct, offering principles and reasoning methods for determining what is right, wrong, just, or unjust. In philosophy, these theories are broadly divided into classical and modern strands, each shaped by the intellectual, cultural, and historical contexts in which they emerged. While classical ethical theories draw heavily on the works of ancient and medieval philosophers and often rest on metaphysical or teleological foundations, modern ethical theories have been influenced by the Enlightenment, scientific developments, and the challenges of pluralistic, technologically advanced societies.

---

[38] Frankena, W. K. (1973). *Ethics* (2nd ed.). (Englewood Cliffs, NJ: Prentice-Hall), pp. 9-11.

Classical ethical thought can be traced to the ancient Greek tradition, where philosophers such as Socrates, Plato, and Aristotle treated ethics as the study of how human beings ought to live in pursuit of the good life. Aristotle's Nicomachean Ethics remains one of the most influential classical accounts, arguing that morality is grounded in the cultivation of virtues stable dispositions to act rightly that enable individuals to achieve eudaimonia, or human flourishing.[39] In this virtue-centered approach, ethical judgment depends not merely on adherence to rules or calculation of outcomes, but on developing the moral character necessary to act appropriately in varied circumstances. Classical virtue ethics thus sees morality as a lifelong practice of self-perfection through reason and habituation.

In the medieval period, classical ethics merged with religious and theological perspectives, most notably in the work of Thomas Aquinas, who integrated Aristotelian philosophy with Christian theology. Aquinas's natural law theory holds that moral principles are derived from human nature and reason, which reflect the eternal law of God .[40] According to this view, moral norms are objective, universal, and accessible through rational reflection, and a morally good act is one that aligns with both human flourishing and divine order.

---

[39] Aristotle, *Op Cit*., pp. 25-27.
[40] Aquinas, T. (1947*). Summa Theologica* (Fathers of the English Dominican Province, Trans.). (New York: Benziger Bros), pp. 47-49.

With the rise of modern philosophy in the seventeenth and eighteenth centuries, ethical theory shifted towards more secular and rationalist foundations. Immanuel Kant's deontological ethics is a hallmark of this period, proposing that morality is grounded in duty rather than consequences. Kant's categorical imperative commands that one act only on principles that could be universally adopted, treating humanity always as an end and never merely as a means.[41] This modern approach emphasizes the autonomy of moral agents and the intrinsic worth of persons, offering a framework for rights-based ethics that remains influential in contemporary human rights discourse.

Another major modern development is utilitarianism, articulated by Jeremy Bentham and refined by John Stuart Mill. Utilitarian ethics evaluates the morality of actions based on their consequences, specifically their capacity to maximize overall happiness or utility .[42] This consequentialist perspective represented a shift from virtue and duty toward measurable social outcomes, aligning with the Enlightenment's empirical and reformist spirit. Utilitarianism has been widely applied in policy-making, economics, and law, though it has also faced criticism for potentially sacrificing individual rights for collective benefit.

In the twentieth century, modern ethical theory expanded to include existentialist and relativist perspectives, which questioned the existence of universal moral truths. Philosophers such as Jean-Paul Sartre argued that moral values are created through

[41] Kant, I. *Op Cit.,* pp. 22-25.
[42] Mill, J. S. *Op Cit*., pp. 14-16.

human freedom and choice rather than discovered as objective facts.[43] Similarly, postmodern thinkers have challenged the universality claimed by earlier theories, emphasizing the role of cultural context, power structures, and language in shaping moral norms.

In the Nigerian context, scholars such as Segun Gbadegesin have argued for an ethical orientation grounded in indigenous values like Omoluabi, which blends personal integrity with communal welfare[44]

The distinction between classical and modern ethical theories is therefore not a matter of superiority, but of orientation. Classical theories tend to focus on the cultivation of virtue, the fulfillment of natural or divine purposes, and the integration of moral life within a cosmic or teleological order. Modern theories, while not abandoning these concerns entirely, often prioritize individual autonomy, rational justification, empirical outcomes, and adaptability to complex social realities. Both traditions contribute valuable insights to moral philosophy, and in practice, many contemporary ethical discussions draw on elements from both. For instance, debates on bioethics, environmental responsibility, and artificial intelligence often combine virtue-based reasoning with rights-oriented duties and utilitarian cost benefit analysis.

---

[43] Sartre, J.-P. (1943). *Being and Nothingness* (H. E. Barnes, Trans.). (New York: Philosophical Library), pp. 555-558.

[44] Gbadegesin, S. *Op Cit*., 65-68.

Ultimately, the enduring relevance of both classical and modern ethical theories lies in their shared aim: to provide coherent, justifiable, and actionable guidance for human conduct. While they differ in their foundational assumptions and evaluative methods, they each affirm that morality is not arbitrary, but rooted in reasoned reflection about the good life, justice, and the responsibilities that bind us to one another.

CHAPTER THREE

ETHICAL DILEMMA IN THE USE OF ARTIFICIAL INTELLIGENCE FOR MORAL DECISION MAKING

3.1 ARTIFICIAL INTELLIGENCE: MEANING AND THEORIES

The juxtaposition of “Artificial” and “Intelligence” initially seems intimidating. The idea of an intelligence not naturally derived evokes apprehension. Popular culture has exerted a significant influence in shaping societal perceptions of AI. Films such as Terminator, I-Robot, and WALL-E portray entities with extraordinary intelligence and the potential to surpass human abilities. While these depictions may seem exaggerated, they underscore underlying concerns regarding the implications of technological progress.

Upon closer examination, the terms “Artificial” and “Intelligence” prove to be complex and multifaceted, contributing to the challenge of defining AI. A thorough conceptual analysis unveils that "artificial" denotes something contrived by humans to mimic or reproduce natural occurrences, while "intelligence" encompasses the capacity to acquire knowledge, solve problems, and adapt to novel situations. Philosopher and Historian of Technology, Jack Copeland defined intelligence as “the ability to adapt one's behaviour to fit new circumstances.”[45] and highlights the “components of intelligence: learning, reasoning, problem-solving, perception, and language-understanding.”[46]

---

[45] Copeland, J. (2000), "What is Artificial Intelligence?" *Alan Turing.Net*, © Copyright B.J. Copeland.

[46] *Ibid*.

Notably, the objectives of artificial intelligence, as articulated by Rochester N in the research proposal on the subject, extend beyond mere replication of human cognition. Rochester underscores the "aspiration for machines to autonomously develop and emulate natural phenomena, facilitated by a semblance of intuition. This quest for intuitive problem-solving capabilities aims to reduce the reliance on human intervention in programming tasks, potentially streamlining problem-solving processes."[47] Artificial Intelligence, thus conceived, represents computer systems that bear a resemblance to the human mind in certain aspects.

Despite the inherent challenges in defining AI, it is discernible that every operational definition of AI constitutes an "abstraction" of human cognition. This abstraction aims to encapsulate aspects of the human mind from diverse viewpoints or levels of abstraction, guided by the conviction that it epitomizes intelligence.[48] Various frameworks for conceptualizing AI exist, each offering distinct perspectives on intelligence emulation. Whether AI is defined in terms of structural similarity to human cognition, behavioral problem-solving capabilities, functional association with human cognitive faculties, or principled decision-making rationality categorized as Structure AI, Behavior AI, Function

---

[47] McCarthy, J. et.al. (1995), *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, (New Jersey: Dartmouth University), p.7

[48] Gabriel, I. (2020), "Artificial Intelligence, Value and Alignment." *Minds and Machines*, vol. 30, pp. 411-437.

AI, and Principle AI respectively[49]. Artificial Intelligence invariably pertains to the endeavor of replicating and augmenting human-like intelligence. It embodies a diverse array of computational systems aimed at mimicking human intelligence.

The goals of artificial intelligence (AI) have evolved, reflecting shifts in the research community's focus and priorities. Initially, AI research aimed to create computers comparable to the human mind, exemplified by ambitious projects like the General Problem Solver that sought "to construct computer programs that can solve problems requiring intelligence and adaptation, and to discover which varieties of these programs can be matched to data on human problem-solving."[50] and Fifth-Generation Computer Systems. However, these aspirations faced skepticism and criticism, leading to a shift towards more practical tasks and individual cognitive functions. This shift marked the transition of AI from a "grand dream" to a more grounded scientific pursuit, emphasizing rigorous methodologies and real-world applications.[51]

While some researchers remain committed to achieving "Human-level AI" or "Artificial General Intelligence" (AGI), mainstream AI has largely focused on developing specialized solutions rather than pursuing general intelligence. Despite divergent

---

[49] Liao S. M., eds, (2020), *Ethics of Artificial Intelligence*, (Oxford: Oxford University Press), p. 42.

[50] Newell, A. Shaw, J. C. & Simon, H. A. (1958), "Report on a General Problem-Solving Program." in *Carnegie Institute of Technology*, (Chicago: University of Illinois), p. 21.

[51] *Ibid*., p. 38.

approaches, there is ongoing debate within the AI community regarding the feasibility and desirability of higher-level AI concepts such as "singularity" or "superintelligence." Moreover, the field of AI is evolving at a rapid pace, where behaviors, once deemed "intelligent" in machines just five years ago, are now scarcely noteworthy. This phenomenon has led to the categorization of Artificial Intelligence into distinct stages, the foremost being Weak AI. Weak AI encompasses the application of AI to address specific problems and technologies. Within Weak AI lies Artificial Narrow Intelligence (ANI), characterized as below human-level AI. ANI operates within confined domains, excelling in specialized areas yet lacking the autonomy to autonomously tackle challenges beyond its defined scope. Nevertheless, ANI often outperforms or equals human performance within its narrow confines. An example of Weak AI includes chatbots and virtual assistants like Siri and Google Assistant, which are programmed to perform specific tasks such as answering questions or providing recommendations within a limited domain. Progressing beyond Weak AI lies the realm of Artificial General Intelligence (AGI), denoted as Strong or Human-Level AI. AGI extends AI's reach across multiple domains, empowering it to autonomously address problems beyond its initial specialization. In numerous areas, AGI achieves parity or superiority to human performance, marking a significant milestone in AI's evolution. A key example of AGI includes OpenAI’s Chat GPT-3 (Generative Pretrained Transformer 3), a language model capable of understanding and generating human-like texts on a wide range of topics, comprehending and responding to various prompts without task-specific programming.

The term "superintelligence" often refers to a hypothetical future stage of AI development where machines possess intelligence far surpassing that of humans in every aspect. For example, while Siri, under Weak AI/ANI, performs tasks that are at par with or below human-level intelligence, under AGI, it may upgrade to perform tasks like writing skills, voice recognition, and even coffee preparation (as a humanoid robot). The goal of Artificial Super Intelligence is for Siri to develop consciousness and awareness that equips it with super-human intelligence and capabilities, such as "solving complex mathematical problems instantaneously or writing a bestseller in a heart (or clock) beat."[52] Superintelligence, in the truest sense of the term, where AI surpasses human intelligence across all domains, has not yet been achieved. While there are advanced AI systems that excel in specific tasks and domains, such as AlphaZero in chess, they do not possess the breadth of general intelligence that would qualify them as superintelligent.

In addition to the stages of AI development, it's essential to explore the foundational subfields that contribute to AI's capabilities. Machine learning, a subset of AI, focuses on enabling machines to learn from data without being explicitly programmed. It encompasses various techniques such as supervised learning, unsupervised learning, and reinforcement learning, which allow AI systems to improve their performance over time through experience. Deep learning, another critical subfield, is inspired by the structure and function of the human brain's neural networks. It involves training artificial neural

[52] Stanford Encyclopedia of Philosophy. (2020). *Ethics of Artificial Intelligence and Robotics.*

networks with large amounts of data to recognize patterns and make predictions. These subfields, along with others like natural language processing and computer vision, constitute the diverse toolkit of AI, enabling advancements in areas such as speech recognition, image classification, and autonomous driving Despite the inherent challenges in delineating its boundaries, AI endeavors to capture and emulate aspects of human cognition, thereby augmenting problem-solving capabilities and facilitating autonomous decision-making processes.

Artificial Intelligence (AI) refers to the simulation of human intelligence processes by machines, particularly computer systems. These processes include learning, reasoning, and self-correction, which are typically associated with human cognition. Philosophically, AI raises questions about the nature of intelligence, consciousness, and the ethical implications of creating systems capable of autonomous decision-making. John McCarthy, who coined the term "Artificial Intelligence" in 1956, defined it as "the science and engineering of making intelligent machines".[53] This definition suggests that AI is not merely about imitating human intelligence but also about constructing systems capable of performing tasks traditionally requiring human intellect. These tasks include problem-solving, decision-making, language understanding, and visual perception.

According to Yann LeCunn, Chief AI Scientist at Meta, Artificial Intelligence is “any information technology capable of solving complex problems that would normally be

[53] McCarthy, J. (2007). *What is Artificial Intelligence?*. Retrieved from http://jmc.stanford.edu/articles/whatisai.html

attributed to humans and animals".[54] Some of these complex problems include perceiving, reasoning, and acting. Artificial Intelligence shares an "intimate and reciprocal relationship with Philosophy".[55] According to the International Encyclopedia of Philosophy, Artificial intelligence (AI) is the possession of intelligence, or the exercise of thought, by machines such as computers. AI's meaning intertwines with longstanding debates on the mind-body problem and the nature of consciousness. René Descartes' dualism, which separates mind and body as distinct substances, challenges the notion of machines possessing true intelligence. According to Descartes, human intelligence is tied to the immaterial mind, which machines, being purely physical entities, cannot emulate. This raises the philosophical question: Can AI exhibit intelligence comparable to that of humans, or is it merely simulating human thought processes without understanding? The philosophical clarification of AI also involves the distinction between "weak AI" and "strong AI." Weak AI refers to systems designed to perform specific tasks and simulate human behavior without true understanding or consciousness. For example, virtual assistants like Siri or Alexa operate under weak AI paradigms. In contrast, strong AI envisions machines with general intelligence and the ability to reason, learn, and understand as humans do. John Searle critiques the possibility of strong AI in his famous

---

[54] Le Cun, Y. Destination AI: Introduction to Artificial Intelligence." Open Classrooms, openclassrooms.com/en/courses/7078811-destination-ai-introduction-to-artificial-intelligence. Accessed 13-08-2025.

[55] Brighton, H. (2012), *Introducing Artificial Intelligence: A Graphic Guide,* (New York: Icon Books Limited), p. 21.

"Chinese Room" argument, asserting that a machine can process symbols and appear intelligent without genuine understanding[56]

Computationally, modern AI theories focus on machine learning and neural networks. Machine learning enables systems to improve performance through data exposure rather than relying on pre-defined rules. Within this framework, supervised learning, unsupervised learning, and reinforcement learning represent core approaches[57]. Neural network theory, inspired by the human brain's structure, is a cornerstone of contemporary AI. Deep learning, which uses multi-layered neural networks, has enabled breakthroughs in image recognition, natural language processing, and autonomous driving[58]. Unlike symbolic AI, connectionist theories emphasize distributed processing, where knowledge emerges from the interaction of numerous simple units rather than explicit symbolic rules.

Probabilistic and Bayesian approaches also form an important theoretical foundation for AI. These theories view intelligence as the capacity to make rational predictions and decisions under conditions of uncertainty. By applying probability theory and statistical reasoning, Bayesian models allow machines to weigh evidence and update beliefs dynamically. These approaches have proven particularly effective in real-world domains,

---

[56] Searle, J. (1980), "Minds, Brains, and Programs." *Behavioral and Brain Sciences*, 3(3), 417-457.

[57] Mitchell, T. (1997), *Machine Learning*, (New York: McGraw-Hill), p. 67.

[58] LeCun, Y. Bengio, Y. & Hinton, G. (2015), *Deep learning*. Nature, 521(7553), 436–444.

such as robotics, diagnostic systems, and speech recognition, where uncertainty is inevitable. The meaning of AI encompasses both the simulation and augmentation of human intelligence through machines. Its theories span philosophical debates about the mind, symbolic logic-based approaches, cognitive models, and computational frameworks like neural networks and probabilistic reasoning. While early theories emphasized symbolic reasoning, contemporary AI largely revolves around adaptive, data-driven models that mirror human learning processes. Each theoretical strand, from Turing's behavioral test to Bayesian inference models, has contributed to shaping AI into a field that is both scientifically rigorous and technologically transformative.

3.2 THE ROLE OF ARTIFICIAL INTELLIGENCE IN MORAL DECISION MAKING

Artificial Intelligence (AI) is increasingly being integrated into domains that involve moral and ethical decision-making, raising significant questions about its role, capacity, and limitations in such contexts. Traditionally, moral decisions were viewed as uniquely human, guided by conscience, values, empathy, and cultural norms. However, as AI systems are deployed in areas such as healthcare, criminal justice, autonomous vehicles, and military operations, they are confronted with situations that require weighing competing moral values and making choices that affect human well-being. This has led to a growing field of research in "machine ethics," which investigates how moral reasoning can be embedded into AI systems.

One of the key roles of AI in moral decision-making is its ability to process vast amounts of data and provide objective analysis in ethically complex scenarios. For instance, in medical contexts, AI systems can assist doctors in prioritizing patients for organ transplants by evaluating survival probabilities, quality of life outcomes, and fairness across demographics. In such cases, AI can reduce human bias and offer evidence-based recommendations, which can enhance fairness and consistency[59]. However, the challenge arises when these decisions involve subjective moral values, such as whether to prioritize the young over the elderly or to value individual rights over collective good.

AI also plays a significant role in autonomous systems, such as self-driving cars. These machines must make rapid moral decisions in life-and-death situations, often referred to as the "trolley problem" of AI. For example, if an accident is unavoidable, should the car prioritize the safety of its passengers or pedestrians? Studies like the Moral Machine experiment[60] highlight the difficulty of programming AI with universal moral standards, as cultural and societal values differ widely. This demonstrates that while AI can be programmed with ethical guidelines, it cannot easily reconcile the diversity of human moral perspectives.

---

[59] Russell, S. & Norvig, P. (2020), *Artificial Intelligence: A Modern Approach* (4th ed.). (London: Pearson Publishing), p. 77.

[60] Awad, E. Dsouza, S. Kim, R. Schulz, J. Henrich, Shariff, J. A. Bonnefon, J. F. & Rahwan, I. (2018), The Moral Machine experiment. *Nature*, 563(7729), 2018, 59–64.

Another role of AI in moral decision-making is in law enforcement and criminal justice. Predictive policing algorithms and risk assessment tools are increasingly being used to determine sentencing or parole outcomes. Ideally, these systems should enhance fairness by relying on data rather than subjective judgment. However, evidence shows that biased data can lead to discriminatory outcomes, reproducing and even amplifying existing social inequalities [61]. This raises ethical concerns about accountability: if an AI system makes an unjust decision, who bears moral responsibility—the developer, the user, or the machine itself? Theoretically, approaches to embedding morality into AI draw from philosophical traditions. Utilitarian models emphasize maximizing overall happiness or minimizing harm, which can be encoded as optimization problems. Deontological approaches, inspired by Kant, suggest that AI should follow strict ethical rules regardless of outcomes. Virtue ethics, though harder to implement, encourages designing AI systems that align with human values, empathy, and character[62]. These approaches highlight that moral decision-making in AI is not merely technical but fundamentally philosophical, requiring collaboration between engineers, ethicists, and policymakers. AI's role in moral decision-making is both promising and problematic. It can enhance fairness, reduce bias, and process ethical dilemmas more systematically than humans in some cases. Yet, it also faces profound challenges, including cultural relativism, the risk of algorithmic bias, and questions of accountability. Ultimately, AI cannot replace human moral agency but can

[61] T. Mitchell, *Op Cit.,* p. 71.

[62] Kant, I. (1997), *Groundwork of the Metaphysics Morals*, trans. Mary Gregor. (Cambridge: Cambridge University Press), pp. 36–49.

serve as a tool to support ethical reasoning, provided it is designed and governed responsibly. The future of moral AI will depend on balancing technological capabilities with human values, ensuring that machines serve humanity rather than replace its ethical judgment.

3.3 SOME ETHICAL ISSUES INVOLVED IN THE USE OF ARTIFICIAL INTELLIGENCE

The growing integration of Artificial Intelligence (AI) into daily life has opened new possibilities but also sparked a range of ethical concerns. One of the most widely discussed issues is bias and discrimination. Since AI systems learn from data, they often reproduce existing social inequalities embedded in that data. For instance, facial recognition technologies have demonstrated higher error rates in identifying women and people of color, raising concerns about racial profiling and unfair treatment[63]. Similarly, hiring algorithms trained on historical company data may perpetuate gender or racial biases, disadvantaging minority groups in employment opportunities. This problem challenges the assumption of neutrality in AI and emphasizes the need for fairness and accountability in algorithmic design.

---

[63] Aristotle, (1999), *Nicomachean Ethics*, trans. Terence Irwin, (London: Hackett Publishing), pp. 28–35.

Another pressing ethical concern is privacy and surveillance. AI thrives on access to massive datasets, often containing sensitive personal information. The widespread use of AI-powered surveillance systems by corporations and governments has sparked debates about the erosion of privacy and the potential misuse of data. In some cases, governments employ AI-driven technologies for mass surveillance and social control, raising questions about the balance between security and individual freedom[64]. Ethical concerns also arise when companies collect and exploit user data without informed consent, highlighting the need for stronger data protection regulations. Accountability and responsibility form another key ethical issue. Unlike traditional tools, AI systems can operate autonomously, making it difficult to assign blame when errors occur. For example, if a self-driving car causes an accident, determining whether responsibility lies with the developer, manufacturer, user, or the AI system itself becomes complex. This accountability gap not only challenges legal frameworks but also raises moral questions about human reliance on autonomous decision-making systems.

The impact of AI on employment and economic inequality is also ethically significant. Automation threatens to displace millions of workers, especially in manufacturing, customer service, and transport industries. While AI can increase efficiency and productivity, it risks widening economic inequality between those with access to

[64] Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, (New York: PublicAffairs), p. 54.

advanced technological skills and those without[65]. This raises ethical questions about social justice, economic redistribution, and the responsibility of governments and corporations in ensuring fair transitions for displaced workers. Another issue concerns human autonomy and control. As AI systems become more advanced, there is a danger that humans may relinquish too much decision-making power to machines. In areas such as healthcare, finance, and military operations, over-reliance on AI could erode human judgment and accountability. For example, the development of autonomous weapons raises deep ethical concerns about whether machines should ever be entrusted with life-and-death decisions. The loss of human agency in such scenarios underscores the importance of keeping humans in the decision-making loop.

AI raises questions of ethical alignment and cultural diversity. Since most AI systems are designed by a handful of global corporations, they often reflect the values and cultural norms of their creators. This can result in "value imposition," where Western ethical frameworks dominate global applications of AI, disregarding diverse moral traditions.[66] Aligning AI with pluralistic human values remains a significant challenge, requiring cross-cultural dialogue and inclusive policymaking. In sum, the ethical issues surrounding AI are multifaceted and interconnected. Bias, privacy violations, accountability gaps, economic inequality, loss of autonomy, and cultural alignment all

---

[65] Brynjolfsson, E. & McAfee, A. (2014), *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, (New York: W.W. Norton), p. 32.

[66] Liao, S. M. *Op. Cit*, p. 56

reveal that AI is not merely a technical challenge but a deeply moral one. Addressing these issues requires interdisciplinary collaboration between computer scientists, ethicists, policymakers, and civil society to ensure that AI develops in ways that uphold human dignity, justice, and fairness.

3.4 PROBLEMS ASSOCIATED WITH THE USE OF ARTIFICIAL INTELLIGENCE IN MORAL DECISION MAKING

One of the major problems associated with the use of Artificial Intelligence (AI) in moral decision-making lies in the lack of genuine moral reasoning and consciousness. Unlike humans, AI systems do not possess moral intuitions, emotions, or a sense of empathy, which are often central to ethical decision making. For instance, while a human judge may consider mercy, compassion, and fairness beyond strict legal codes, an AI programmed to follow data-driven logic may rigidly enforce rules without considering broader human values[67]. This inability to integrate human emotions into decisions makes AI decisions appear detached, mechanical, and potentially harmful when applied to moral contexts.

Another significant issue is bias in algorithms, which often stems from the datasets used to train AI systems. If the data fed into AI models contains social prejudices, historical inequalities, or systemic discrimination, the AI will reproduce and reinforce those biases

[67] Coeckelbergh, M. (2020), *AI Ethics*, (London: MIT Press), p. 87.

in its decisions. For example, research has shown that facial recognition technologies tend to misidentify people of darker skin tones more frequently, leading to unfair outcomes in areas like policing and security. In moral decision making, such biases could result in discriminatory judgments that disproportionately harm marginalized groups. Thus, instead of offering neutral solutions, AI might amplify ethical problems already existing in society.

The problem of accountability also poses a major challenge. When AI makes a morally questionable or harmful decision, it is often unclear who should be held responsible, the programmer, the organization deploying it, or the AI system itself. This "responsibility gap"[68] creates serious ethical and legal dilemmas, especially in life-or-death situations such as autonomous weapons or self-driving cars. For instance, if a self-driving car decides to swerve in a way that causes one person's death instead of another's, determining who is morally or legally responsible for that choice becomes highly problematic. The absence of a clear framework for accountability undermines trust in AI-driven moral decision making.

Additionally, lack of transparency and explainability in AI systems raises ethical concerns. Many AI systems, especially those using deep learning, operate as "black boxes," meaning their internal reasoning processes are not easily understandable by humans.

---

[68] Wallach, W. & Allen, C. (2009), *Moral Machines: Teaching Robots Right from Wrong*, (Oxford: Oxford University Press), p. 77.

When an AI makes a moral decision, stakeholders may find it difficult to trace why it made such a choice. This lack of explainability undermines trust and prevents ethical scrutiny, as decisions affecting human lives should be subject to reasoning that can be understood and questioned. Without transparency, AI may be perceived as arbitrary, even if it technically follows its programmed logic.

Another problem is the overreliance on utilitarian logic by many AI models. In moral decision-making scenarios, AI systems tend to weigh options based on maximizing benefits or minimizing harm, often adopting a cost-benefit analysis approach. However, not all moral dilemmas can be reduced to utilitarian calculations. Ethical theories such as deontology emphasize duties, rights, and principles that should not be violated, regardless of the consequences. For example, sacrificing one person to save five may seem rational from a utilitarian perspective, but deontologists argue that it violates the moral duty not to kill an innocent person. AI's inability to balance competing ethical frameworks leads to moral oversimplification and potentially unjust decisions.[69]

AI systems face the challenge of contextual understanding. Moral decisions are often shaped by culture, tradition, religion, and human experiences. A decision that seems morally acceptable in one culture may be offensive or unethical in another. AI, however, lacks the ability to appreciate such cultural nuances and may make decisions that are contextually inappropriate or even harmful. In medical decision making, an AI might

[69] James, R. (2012), *The Elements of Moral Philosophy*, 7th ed., (New York: Mc Graw-Hill Education), p. 25.

recommend life-saving surgery without recognizing the cultural or religious prohibitions against certain medical interventions. This lack of contextual moral sensitivity undermines the legitimacy of AI in real-world ethical dilemmas. The risk of dehumanization is another pressing concern. If societies increasingly delegate moral decisions to AI systems, there is a danger that human beings will lose their sense of moral responsibility and agency. By outsourcing ethical judgments to machines, humans may gradually abdicate their own moral capacities, relying instead on automated systems to decide what is right or wrong. This could erode moral growth, human dignity, and the social processes through which humans learn, debate, and refine their ethical values. Such reliance may reduce human beings to passive agents, subject to the dictates of machines that cannot truly grasp the essence of human morality.

CHAPTER FOUR

EVALUATION, RECOMMENDATIONS AND CONCLUSION

4.1 EVALUATION

The use of Artificial Intelligence (AI) in moral decision-making brings to light several ethical dilemmas that cut across culture, law, philosophy, and technology. These dilemmas highlight the challenges of entrusting machines with tasks that are deeply human, value-laden, and context-sensitive. Evaluating these dilemmas reveals the limitations of AI in replicating human moral reasoning and underscores the risks associated with its deployment in sensitive domains. One major dilemma lies in the diversity of moral norms across societies. Human morality is not universal; it is shaped by culture, religion, and social context. The MIT *Moral Machine* experiment revealed striking differences in how people across cultures evaluated moral tradeoffs, such as

whether to save younger versus older individuals in autonomous vehicle accidents[70]. This variability makes it ethically problematic to design AI systems with a one-size-fits-all moral code. If AI is programmed using Western ethical frameworks, for example, it may conflict with African, Asian, or Middle Eastern moral traditions. Thus, AI risks becoming an instrument of cultural bias rather than a neutral tool.

Stuart Russell identifies the "value alignment" problem as central to AI ethics. AI systems must operate with objectives that align with human values.[71] Yet, specifying these values in computational form is nearly impossible, as human morality is often inconsistent and situational. For instance, a utilitarian algorithm may prioritize saving the greatest number of lives in a crisis, while a deontological framework may forbid sacrificing one person for others, even if it maximizes outcomes. Embedding either principle into AI excludes the other, creating an ethical dilemma about which moral system should prevail. This misalignment shows that AI cannot avoid normative commitments; designers must decide which values to encode, raising concerns about whose values dominate.

Another pressing dilemma is accountability. AI decision-making introduces what is often called the "responsibility gap." If an autonomous drone mistakenly targets civilians or an algorithm denies life-saving treatment, it is unclear who should be held responsible — the

---

[70] Awad, E., et al. (2018). *The Moral Machine Experiment*. *Nature*, 563(7729), 59–64.

[71] Russell, S. J. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. (London: Viking Press), p. 56.

programmer, the deploying institution, or the AI itself[72]. Unlike human agents, AI lacks moral agency and cannot be punished or held accountable in a meaningful way. This creates a moral hazard: companies and governments might evade responsibility by attributing harmful outcomes to "the algorithm." Such accountability gaps undermine justice and human rights protections.

AI systems are only as fair as the data they are trained on. Studies show that algorithms used in facial recognition, predictive policing, and healthcare reproduce existing social inequalities[73]. When these systems are tasked with moral decision-making, the risk of amplifying discrimination increases. For example, an AI trained predominantly on Western data may misinterpret cultural contexts in Africa or Asia, producing morally flawed decisions. The dilemma is that AI is presented as an objective decision-maker, yet in reality, it inherits human biases and may systematically disadvantage vulnerable groups. This raises ethical questions about justice, fairness, and equality in AI-guided moral choices. AI often functions as a "black box," where even developers cannot fully explain how decisions are reached. This opacity undermines transparency, especially when moral decisions are involved. If an AI denies parole to an inmate or decides whom to prioritize in medical triage, affected individuals deserve to know the reasoning behind

---

[72] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. IEEE.

[73] Afreen, J., et al. (2025). *Systematic Literature Review on Bias Mitigation in AI*. (London: Springer), p. 23.

those decisions[74]. The lack of explainability not only erodes public trust but also challenges democratic accountability. People are more likely to accept morally difficult outcomes when they understand the reasoning; AI deprives them of this possibility, making its decisions appear arbitrary or unfair.

An additional dilemma concerns the potential erosion of human dignity and moral autonomy. Delegating moral choices to machines risks treating individuals as data points rather than moral agents with intrinsic worth. For example, in healthcare, patients may feel dehumanized if AI systems, rather than doctors, decide who receives treatment. This challenges Kantian ethics, which emphasizes respecting individuals as ends in themselves rather than means to an end[75]. By outsourcing moral reasoning, society risks undermining the very concept of human dignity.

The ethical dilemmas associated with AI in moral decision-making reveal that the problem is not simply technological but deeply philosophical and political. Cultural relativism shows the difficulty of universal moral coding, the value alignment problem exposes normative biases in design, and the responsibility gap questions legal accountability. Furthermore, bias and transparency issues demonstrate that AI may reproduce injustice under the guise of objectivity, while concerns about human dignity highlight the existential risks of ceding moral authority to machines. Ultimately, AI cannot resolve moral disagreements; rather, it magnifies them by embedding contested

[74] OECD. (2019). *OECD Principles on Artificial Intelligence*. OECD Publishing.

[75] Russell, S. J. *Op. Cit*., p. 78.

values into decision-making processes. This evaluation demonstrates that while AI may support moral deliberation, it should not replace human judgment in contexts that involve fundamental ethical tradeoffs. Instead, a hybrid model where AI assists but humans retain moral and legal accountability appears to be the most ethically defensible path forward.

4.2 RECOMMENDATIONS

The evaluation of ethical dilemmas in the use of Artificial Intelligence (AI) for moral decision-making highlights the urgent need for balanced approaches that integrate technological innovation with ethical safeguards. Based on the issues identified, several recommendations are proposed.

First, it is important to ensure human oversight in morally sensitive decisions. AI systems should not function as independent moral agents but rather as tools that support human decision-making. Critical moral judgments such as medical triage, criminal sentencing, or military actions, must remain under the authority of accountable human actors. This approach preserves moral responsibility while using AI's efficiency to enhance decision-making processes.

Second, there should be a pluralistic and participatory approach to AI design. Since morality varies across cultures, AI systems must be designed with input from diverse stakeholders, including ethicists, policymakers, religious leaders, cultural representatives, and affected communities. This ensures that no single moral framework dominates, and it

reduces the risk of cultural bias. Multistakeholder engagement also enhances public trust in AI systems by making their design and values more transparent.

Third, legal and regulatory frameworks must be strengthened to close the accountability gap. Governments and international bodies should establish clear rules about who is responsible when AI systems make harmful or unethical decisions. Developers, institutions, and deploying agencies should be held legally accountable, ensuring victims can seek redress. Strong regulation also discourages the misuse of AI and compels organizations to prioritize ethical compliance.

Fourth, efforts should focus on bias detection and mitigation in AI systems. Since algorithms often inherit prejudices from training data, regular auditing, bias testing, and fairness monitoring should become mandatory. Developers should implement technical solutions such as diverse datasets, fairness-aware algorithms, and transparent reporting of limitations to reduce discriminatory outcomes.

Fifth, enhancing transparency and explainability is crucial. AI systems making moral decisions should be designed to provide understandable justifications for their outcomes. This will help affected individuals and institutions to evaluate decisions critically, contest unfair outcomes, and maintain trust in AI. Explainable AI is particularly important in high-stakes domains like healthcare, law, and autonomous vehicles, where decisions have direct human consequences. Finally, the protection of human dignity and moral autonomy must remain central in all AI applications. While AI can assist in processing complex data,

it should never undermine the intrinsic worth of individuals or treat them as mere statistics. Human-centered design approaches that prioritize empathy, fairness, and respect for persons should guide all AI deployments.

4.3 CONCLUSION

The ethical dilemmas surrounding the use of Artificial Intelligence for moral decision-making underscore the complex intersection between technology, philosophy, and human values. While AI has the capacity to enhance decision-making through speed, consistency, and data processing, it also faces significant challenges in handling the moral nuances that define human judgment. Issues such as cultural relativism, value alignment, accountability gaps, bias, and transparency reveal that AI cannot serve as an independent moral authority. Rather, it functions best as a tool that supports, but does not replace, human moral reasoning.

This project has shown that the risks of delegating moral responsibility to machines are substantial. By embedding contested ethical frameworks into algorithms, AI systems can perpetuate cultural bias, amplify injustice, and weaken trust in technological systems. Moreover, the absence of clear accountability raises concerns about justice and responsibility when AI makes harmful decisions. These concerns highlight the need for careful regulation, ethical design, and human oversight in morally sensitive domains.

Despite the challenges, the integration of AI into moral decision-making is not inherently negative. With proper safeguards, AI can support human decision-makers by offering analytical clarity and consistency. However, the preservation of human dignity, fairness, and moral autonomy must remain central. The future of AI in moral contexts depends on striking a careful balance: embracing the benefits of technological advancement while safeguarding ethical principles that define our shared humanity

BIBLIOGRAPHY

Afreen, J., et al. (2025). *Systematic Literature Review on Bias Mitigation in AI*, London: Springer.

Aquinas, T. (1947*). Summa Theologica* (Fathers of the English Dominican Province, Trans.). New York: Benziger Bros, Original work published c. 1265–1274.

Aristotle, (1999), *Nicomachean Ethics*, trans. Terence Irwin, London: Hackett Publishing.

Awad, E. Dsouza, S. Kim, R. Schulz, J. Henrich, Shariff, J. A. Bonnefon, J. F. & Rahwan, I. (2018), The Moral Machine experiment. *Nature*, 563(7729).

Awad, E., et al. (2018). *The Moral Machine Experiment*. *Nature*, 563(7729).

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of Biomedical Ethics* (8th ed.). Oxford: Oxford University Press.

Brighton, H. (2012), *Introducing Artificial Intelligence: A Graphic Guide,* New York: Icon Books Limited.

Brynjolfsson, E. & McAfee, A. (2014), *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, New York: W.W. Norton.

Coeckelbergh, M. (2020), *AI Ethics*, London: MIT Press.

Copeland, J. (2000), "What is Artificial Intelligence?" *Alan Turing.Net*, © Copyright B.J. Copeland.

Frankena, W. K. (1973). *Ethics* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Gabriel, I. (2020), "Artificial Intelligence, Value and Alignment." *Minds and Machines*, vol. 30.

Gbadegesin, S. (1991). *African Philosophy: Traditional Yoruba Philosophy and Contemporary African Realities*. Chicago: Gateway Press.

Gyekye, K. (1997). *Tradition and Modernity: Philosophical Reflections on the African Experience*. Oxford: Oxford University Press.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. IEEE.

James, R. (2012), *The Elements of Moral Philosophy*, 7th ed., New York: Mc Graw-Hill Education.

Kant, I. (1997), *Groundwork of the Metaphysics Morals*, trans. Mary Gregor. Cambridge: Cambridge University Press.

Le Cun, Y. Destination AI: Introduction to Artificial Intelligence." Open Classrooms, openclassrooms.com/en/courses/7078811-destination-ai-introduction-to-artificial-intelligence. Accessed 13-08-2025.

LeCun, Y. Bengio, Y. & Hinton, G. (2015), *Deep learning*. Nature, 521(7553), 436–444.

Liao S. M., eds, (2020), *Ethics of Artificial Intelligence*, Oxford: Oxford University Press.

M.A. Akande, Artificial Intelligence and the limits of Epistemic Justification *LASU Journal of Philosophy*.

MacIntyre, A. (2007). *After Virtue* (3rd ed.). London: University of Notre Dame Press.

McCarthy, J. et.al. (1995), *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, New Jersey: Dartmouth University.

Mill, J. S. (1998). *Utilitarianism* (R. Crisp, Ed.). Oxford: Oxford University Press.

Mitchell, T. (1997), *Machine Learning*, New York: McGraw-Hill.

Newell, A. Shaw, J. C. & Simon, H. A. (1958), "Report on a General Problem-Solving Program." in *Carnegie Institute of Technology*, Chicago: University of Illinois.

OECD. (2019). *OECD Principles on Artificial Intelligence*. OECD Publishing.

Oladipo, O. (2006). *Modernity and African Philosophy*, London: Hope Publications.

Oxford Learner's Dictionary.

Rachels, J. (2012). *The Elements of Moral Philosophy*, 7th ed., New York: MC Graw-Hill Education.

Rawls, J. (1971). *A Theory of Justice*, London: Harvard University Press.

Russell, S. & Norvig, P. (2020), *Artificial Intelligence: A Modern Approach* (4th ed.), London: Pearson Publishing.

Russell, S. J. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. London: Viking Press.

Sartre, J.-P. (1943). *Being and Nothingness* (H. E. Barnes, Trans.). New York: Philosophical Library.

Searle, J. (1980),"Minds, Brains, and Programs." *Behavioral and Brain Sciences*, 3, 3.

Singer, P. (2011). *Practical Ethics*, 3rd ed., Cambridge: Cambridge University Press.

Stanford Encyclopedia of Philosophy. (2020). *Ethics of Artificial Intelligence and Robotics.*

McCarthy, J. (2007). *What is Artificial Intelligence?*. Retrieved from http://jmc.stanford.edu/articles/whatisai.html

Stanford Encyclopedia of Philosophy. (2020, April 30). *Ethics of Artificial Intelligence an Robotics*.

Wallach, W. & Allen, C. (2009), *Moral Machines: Teaching Robots Right from Wrong*, Oxford: Oxford University Press.

Williams, B. (1985). *Ethics and the Limits of Philosophy*, London: Harvard University Press.

Rachels, J., & Rachels, S. (2019). *The Elements of Moral Philosophy* (9th ed.). New York: McGraw-Hill Education.

Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York: PublicAffairs.