IJHSS The Interdisciplinary Journal of Human and Social Studies
ISSN: 2820-7424

The Interdisciplinary Journal of Human and Social Studies, Vol.3, Issue 2, 2024
https://hs-studies.com/

Descartes' Legacy in AI Ethics: Reconsidering Cartesian Dualism in Conceptualizing Ethics of AI

Dr. Daisy DUTTA
Assistant Professor (Guest Faculty)
Department of Philosophy, Cotton University
Panbazar, Guwahati, Assam, India
Email: daisydutta@iitkgp.ac.in
ORCID: 0009-0002-0799-8611

Abstract: *In recent years, Artificial Intelligence (AI) has made significant progress, influencing a variety of aspects of our daily existence. Advances in AI technology must be considered with ethical concerns. There are concerns about how autonomous and advanced AI systems may affect human rights, privacy, and society. This led to the creation of AI ethics, a new area that sets standards for responsible AI development and implementation. Descartes' dualism serves as a fundamental framework for resolving inquiries regarding AI consciousness, moral agency, machine consciousness, moral agency, and ethical responsibility in the context of Artificial Intelligence (AI) ethics. In this paper, the relationship between Cartesian dualism and contemporary AI ethics is investigated, with a focus on the ways in which Descartes' philosophical concepts inform current discussions regarding the ethical treatment of AI entities, the nature of machine consciousness, and the implications for human-machine interactions. The ethical implications of AI technology must be addressed in conjunction with these advancements. As AI systems become increasingly autonomous and sophisticated, there have been apprehensions about their potential impact on human rights, privacy, and society. This has resulted in the emergence of a new field, AI ethics, which aims to establish guidelines and principles for the responsible development and deployment of AI systems. In this paper, the applicability of Descartes' concepts to the field of contemporary AI ethics is investigated. Specifically, it examines the ways in which Cartesian Dualism influences our comprehension of AI consciousness, the ethical treatment of AI entities, and the moral obligations of AI developers and users.*

Keywords*: Cartesian dualism, AI ethics; Descartes' mind, AI advancements, ethical guidance*

1. Introduction

In the history of philosophy of mind, Descartes has had an influential impact on the various discussions of contemporary philosophers regarding the issues of mind, intellect, consciousness, mind-body dualism, and so on. The concept of mind in terms of thought or consciousness as proposed in Descartes' philosophy and the relation between substance and its attributes found in Cartesian dualism have built a strong foundation in the modern era of philosophy of mind (Alanen 1989; Cottingham 2008). In the contemporary era, with the proliferation of artificial intelligence (AI), Descartes' philosophy of mind, his dualism in

particular, offers a vital starting point for investigating modern inquiries on AI awareness and its moral capacities. Descartes's dualism of mind and body as two distinct and separate entities laid the foundation for exploring some of the crucial philosophical problems of artificial intelligence and raises doubts about the potential for AI to attain authentic self-awareness. Descartes' philosophy of mind continues to inform contemporary debates on the nature of consciousness and the ethical ramifications of AI (Nath 2010; Blackshaw 2023; Hildt 2023). However, the relationship is not well explored.

This paper argues that the amalgamation of Descartes's philosophy with AI is both intriguing and essential, as it promotes a more profound comprehension of the consequences of today's AI advancements. As researchers and engineers strive to imbue machines with increasingly sophisticated cognitive capabilities, Descartes' ideas prompt us to critically evaluate the potential consequences of these advancements. To adopt a Cartesian viewpoint, if consciousness is inherently connected to immaterial cognition, then artificial intelligence, which is a result of material computing, may lack authentic consciousness (Descartes 1954; Gonzalez 2019). This perspective strengthens a critical position on the question of AI awareness and its moral agency. That is, moral agency necessitates a consciousness that is capable of self-awareness and reason, as implied by Descartes' dualism. If artificial intelligence is determined to be incapable of exhibiting authentic consciousness, it could be contended that it also lacks moral agency. In contrast, if consciousness is seen as an outcome of intricate information processing, artificial intelligence may potentially question this Cartesian paradigm. This standpoint also suggests that artificial intelligence would not be subject to moral assessment analogous to that of humans. However, since the fast progress in artificial intelligence has prompted efforts to develop computers that have cognitive capabilities similar to those of humans, including thinking, learning, and problem-solving, it is also important to consider whether they can or should be subjected to ethical principles. In such case, questions, such as, what are the implications for creating intelligent machines that emulate human cognition? How do we address the ethical and moral concerns arising from the potential development of conscious AI entities, if such a feat is possible? and so on are significant and crucial. This prompts us to reassess whether consciousness is inherently independent of materiality or whether it may arise from sophisticated computational procedures. This philosophical stance introduces a framework that can be applied to evaluate the nature of AI consciousness and moral responsibility.

This paper primarily focuses on exploring the links between Descartes' philosophy of mind and the developing field of artificial intelligence, specifically the moral agency of AI. The paper analyses Descartes' conception of mind vis-a-vis the mind-body relationship of Cartesian dualism and attempts to explore the ethical implications of AI advancements. With an analysis of Descartes' mind-body dualism, this study attempts to understand the importance of considering Cartesian concept of mind and body in AI development and sustainable implementation. This study also develops certain arguments to support that Descartes's non-computational theory of mind has the potential to substantiate the research endeavors that primarily deal with AI ethics.

This paper has four sections. In the introduction, it briefly outlines the rationale of the study and its relevance. In the second section, Descartes' standpoint on the conception of mind and the significance of his mind-body dualism have been explained in two sub-sections separately. Third section comprises of a brief analysis of the modern-day AI advancements, the ethical issues associated with it and the area of AI ethics as a branch of applied ethics. In the fourth section, arguments are offered to support the claim that Cartesian Dualism has significant potential to contribute in ethical guiding to the fundamental philosophical problems of modern-day artificial intelligence. The last section concludes the paper.

2. Descartes's concept of mind and Cartesian Dualism:

Descartes can be considered as the first thinker to thoroughly investigate the implications of the emerging mechanical philosophy on the understanding of human existence. Descartes pondered the possibility of developing a machine that could not be distinguished phenomenologically from a human being. Descartes asserted that human and animal bodies could be comprehended as complex and intricately structured machines (Descartes 1954). Descartes believed that the essence of our material bodies is extension, and thus behavior of bodies is entirely governed by certain mechanical laws. Nevertheless, Descartes maintained that humans, in contrast to non-human animals, are not just physical bodies; they are combinations of material bodies and immaterial souls. In Descartes' philosophy, the conception of mind receives more attention than the conception of body; however, Descartes tried to build a bridge between the mechanical body and the psychological mind with his mind-body interactionism. But, before analyzing Descartes's mind-body dualism, it is important to understand his conception of mind.

2.1 Descartes' Concept of mind:

Descartes initiates his investigation to establish something that is completely certain and beyond any doubt by casting doubt on every conceivable thing. Through this process, he realizes that despite being able to doubt everything, we cannot doubt our act of doubting. Therefore, based on this reasoning, we can infer that there must also be an entity or possessor of this state of doubting, which is a form of thinking. Consequently, although we dismiss every possible thing by doubting it, we cannot refute the fact that we are something that can at least think or doubt. Descartes concludes that he exists, at least as a thinking being, and he views this conclusion as unquestionably true. With this perspective in mind, Descartes has introduced the renowned phrase 'Cogito Ergo Sum' (I think, therefore I am), which he has deemed as the fundamental principle of philosophy. As Descartes writes in *Discourse on Method*:

"*But immediately afterwards I noticed that whilst I thus wished to think all things false, it was absolutely essential that the 'I' who thought this should be somewhat, and remarking that this truth 'I think, therefore I am' was so certain and so assured that all the most extravagant suppositions brought forward by the sceptics were incapable of shaking it, I came to the conclusion that I could receive it without scruple as the first principle of the Philosophy for which I was seeking*." (Descartes 1954).

The key points to be noted from the above definition and the '*Cogito*' argument of Descartes are:

- Descartes affirmed the certainty of his existence from the certainty of his thought and while doing so, he derived his existence only in the first-person perspective. He neither affirmed the existence of any other mind nor any other body nor even his own bodily structure. The 'I' of the statement is totally psychological, not physical.
- His existence is only valid for the time when he is thinking, i.e. his existence itself is not necessary, but if he thinks then it necessarily follows that he exists. Therefore, it is contradictory for Descartes to believe that at a same time when one is thinking, does not exist.
- Descartes did not derive the statement '*Cogito Ergo Sum*' on the basis of some empirical evidence, nor does he derive it by means of deduction. Descartes has clearly stated that his knowledge of the '*Cogito*' is through intuition.

Descartes defined mind in terms of thinking substance. Now, it is important to examine that what Descartes meant by the term 'thinking' when he defined 'mind' as a 'thinking thing'. Thinking, according to Descartes, does not cover only the acts of the reason or the pure intellect, but all kinds of conscious mental states and events, including the acts of the will and the imagination, as well as emotions and sense-

perceptions. What is common to all these different sorts or modes of thinking, is that we are aware of their occurrences while having or entertaining them. That is, they are 'thoughts' in Descartes' sense only in so far as they are consciously perceived. In other words, we can say that according to Descartes, we cannot have any thought of which we are not aware or conscious. According to Descartes, mind is *"A thing that thinks. What is that? A thing that doubts, understands, affirms, denies, is willing, is unwilling, and also imagines and has sensory perceptions."* (Descartes 1951). This implies that Descartes employed the term 'thinking' to designate all mental processes and activities, such as, cognitive, emotional, voluntary and so on.

2.2 Cartesian Dualism:

Cartesian dualism maintains the standpoint that the subjects of our experience are conceived as separate from the physical bodies. But, most importantly, the key point of Cartesian dualism is that though Descartes interprets that the mental states and the bodily states are distinct from each other, yet there is an interaction possible between them. Descartes offers three arguments in favour of the mind-body dualism or often known as Cartesian dualism:

(a) The Argument from Conceivability/Inconceivability: Using the notion of conceivability and inconceivability, Descartes argues that our conception of ourselves as thinking creatures is basic and not dependent on sensory observation. Descartes defines the mind as an independent cognitive entity, distinct from the physical body, which is mainly characterized by its extension. This argument may be more effectively elucidated by referring to Descartes' original statements from *Discourse on Methods*:

"And then, examining attentively that which I was, I saw that I could conceive that I had no body, and that there was no world nor place where I might be; but yet that I could not for all that conceive that I was not. (…) From that I knew that I was a substance the whole essence or nature of which is to think, and that for its existence there is no need of any place, nor does it depend on any material thing; so that this 'me', that is to say, the soul by which I am what I am, is entirely distinct from body, and is even more easy to know than is the latter; and even if body were not, the soul would not cease to be what it is." (Descartes 1954).

This argument posits that although we may dismiss the existence of physical entities that we observe via our senses, we cannot question the existence of our mind, which is accountable for the perception of these sensory experiences. Hence, the mind and body are distinct things.

(b) The Epistemological argument: Descartes' concept of clear and distinct thoughts forms the basis of this argument. Descartes insists on his clear and accurate comprehension of his mind, which is different from the notion of the body. Similarly, he maintains a separate notion of the body that is not reliant on his mind. Descartes analyzes the epistemological argument by contemplating two separately defined concepts. First and foremost, he has an unambiguous and accurate perception of himself as a cognitive being that does not possess any physical extension. Second, he has a clear perception of the body as an elongated thing devoid of cognitive capacity. Therefore, it is undeniable that I am distinct from my corporeal form and capable of existing autonomously from it (Descartes 1951).

Descartes argues that the mind may exist independently, dissociated from a physical body, but the essence of the body cannot be understood without being in a corporeal state. Descartes theorizes that the essence of the 'I' cannot be the same as the nature of the body since he can see himself living without a physical body. When we consider the mind and body as independent things, in accordance with Descartes' concept of clear and distinct ideas, we may emphasize two important features. In the first instance, Descartes claims to have a clear and distinct understanding of the fundamental nature of his mind, while he has a

precise, but separate comprehension of the fundamental nature of the body. However, Descartes has never explicitly asserted that his views on the nature of the body are unambiguous and distinct. He just acknowledges that he has a coherent understanding of his physical form. This implies that although Descartes claims to have a precise understanding of his body, he has never claimed that the understanding of his body is equivalent to his understanding of mind about which he has clear and distinct thoughts. Furthermore, Descartes argues that the validity of every clear and clearly defined experience is guaranteed by the presence of God, who is beyond from any kind of deception.

(c) Argument based on the concept of divisibility or indivisibility: Descartes argues that the mind is intrinsically indivisible since it is both simple and non-spatial, hence it lacks extension. The physical body, being a concrete reality manifesting in physical space, may be philosophically divided. Based on the evident differences between the mind and body, it may be inferred that they are not indistinguishable entities. Descartes considers this issue in his *Meditations on First Philosophy* in the following manner:

"*(…) there is a great difference between the mind and the body, in as much as the body is by its very nature always divisible, while the mind is utterly indivisible. For when I consider the mind, or myself in so far as I am merely a thinking thing, I am unable to distinguish any parts within myself; I understand myself to be something quite single and complete... By contrast, there is no corporeal or extended thing that I can think of which in my thought I cannot easily divide into parts; and this very fact makes me understand that it is divisible. This one argument would be enough to show me that the mind is completely different from the body."* (Descartes 1951).

According to Descartes, he is only a cognitive being or intellect, with the ability to understand, affirm, negate, desire, imagine, and even perceive sensory impulses. Ascription of such attributes to fully extended, non-cognitive things such as stones, tables, pencils, and other physical objects is irrational. Therefore, only the intellect has these states. Conversely, it is irrational to ascribe attributes such as size, shape, quantity, and movement to intangible entities that engage in cognitive processes. For example, the concept of a shape that exhibits no extension is unfathomable. Therefore, the intelligent mind cannot be divided into distinct elements, just as the physical body cannot be regarded as an entity that cannot be divisible. Hence, there are few essential reasons to claim that the mind and body are distinct entities. Nevertheless, a fundamental element of Descartes' comprehension of the mind-body phenomenon is that while he perceives them as distinct entities, he also contends that human nature is shaped by the amalgamation of the mind and body.

This elucidation of Cartesian dualism demonstrates that Descartes' mind is non-mechanistic and non-computational. Simultaneously, it demonstrates that the cognitive faculties, including thinking, imagining, and sensation, cannot be attributed to the body. Consequently, according to Descartes, a machine is incapable of thinking or exhibiting consciousness (Gonzalez 2019). Nevertheless, artificial intelligence is an endeavor to develop intelligent machines that can reason like humans. It is important to mention that Descartes himself entertained the notion of developing an intelligent machine that resembled a human; however, he maintains that such machines are incapable of possessing authentic intelligence or reason. Precisely, Descartes asserted that machines are incapable of attaining the same level of consciousness or reasoning abilities as humans. This raises several concerns regarding the genuine moral status of human-like machines, and the ethical issues with the deployment of AIs.

To analyze, what are the contributions of Cartesian Dualism that led the discussions of AI advancements and the philosophical problems associated with it, and how Descartes' philosophy may have substantial

contributions in the discussions of AI Ethics, first of all, it is better to understand what AI is, what are its advancements, what ethical issues it may raise, and what AI Ethics can do.

3. Artificial Intelligence: AI Advancements, Ethical Issues, and the AI Ethics:

Artificial Intelligence (AI) commonly refers to the technology allows machines to replicate a wide range of intricate human abilities; it is an independent and self-learning entity capable of performing intelligent tasks, such as, acquiring knowledge from experience, logical thinking, problem-solving and so on just like humans (Sheikh et al., 2023). Multiple intellectuals have attempted to articulate the concept of AI from diverse viewpoints. Nevertheless, it remains challenging to provide a widely acknowledged definition of AI (Russell and Norvig 2020).

The field of Artificial Intelligence (AI) has had remarkable progress in recent decades, resulting in substantial enhancements across several industries, such as healthcare, banking, and transportation (Y Lu 2019; Dwivedi et al., 2021). AI is transforming our technological interactions and decision-making processes via advanced machine learning algorithms and autonomous systems. Although these breakthroughs provide considerable advantages, these technological developments also give rise to a multitude of ethical dilemmas that need thorough scrutiny and must be resolved to guarantee responsible and fair use (Zhou and Chen 2019; Gabriel 2020). This section of the paper delves into the key advancements in AI, the fundamental ethical concerns linked to the development of artificial intelligence, analyzing their consequences and suggesting possible remedies.

3.1 Key Advancements in AI

Today we may see significant progress in Artificial Intelligence. Artificial intelligence techniques such as deep learning and neural networks, Natural Language Processing (NLP), Self-governing systems, AI in healthcare, and robotics and automation massively impacted our lives, and makes life easier. Some of the key advancements in AI are:

(a) Deep learning and Neural Networks: A significant breakthrough in AI is the progress made in developing and improving deep learning algorithms and neural networks (Aggarwal 2018). Deep learning is a branch of machine learning that focuses on training artificial neural networks with numerous layers to identify patterns and make choices (Goodfellow et al., 2016). This method has allowed substantial advancements in domains such as, image and voice recognition. Deep learning models are used in many applications such as face recognition systems, voice assistants, and driverless cars (Soori et al., 2023). The success of deep learning may be mostly attributable to the abundance of extensive datasets and the significant advancements in computer capacity (LeCun et al., 2015). Advancements in technology, such as Graphics Processing Units (GPUs) and specialized AI chips, have significantly sped up the training of intricate models, resulting in significant advancements in natural language processing and computer vision (Gupta 2021). For instance, companies like Tesla and Waymo use deep learning in autonomous vehicles to interpret visual data from sensors and cameras.

(b) Natural Language Processing (NLP): It is the field of study that focuses on the interaction between computers and human language. It involves developing algorithms and models that enable computers to understand, interpret, and generate human language in a way that is like how humans do (Brown et al., 2020). The field of Natural Language Processing (NLP) has seen significant progress, especially with the emergence of expansive language models like as OpenAI's GPT-3. These models can comprehend and produce text that resembles human language, hence facilitating the development of applications like chatbots, language translation, and content production (Myers et al., 2024). The progress in natural language processing (NLP) has enhanced the interactions between humans and

computers by simplifying the development of systems capable of comprehending and reacting to inputs in natural language (Vaswani et al., 2017). These developments have ramifications for many areas, including customer service, education, and entertainment. AI-powered content production systems may aid in composing articles, producing marketing materials, and developing imaginative writing. For example, tools like Google Translate, powered by neural machine translation, allow real-time translation of multiple languages. This capability facilitates communication across borders, aiding in international business, education, and travel.

(c) Self-governing systems: The progress of autonomous systems has the capacity to transform companies and enhance efficiency in several areas. Autonomous systems, such as self-driving automobiles and drones, signify a substantial advancement in artificial intelligence capabilities (Goodall 2014). These systems use a fusion of sensors, machine learning algorithms, and real-time data processing to autonomously navigate and execute tasks without human interaction (Chen and Chien 2019). Tesla and Waymo are leading the way in the development of autonomous cars, with the goal of improving road safety, decreasing traffic congestion, and offering new transportation choices. AI-enabled drones are used for several purposes, including product delivery, surveillance, and environmental monitoring (Suryanarayana 2023).

(d) Artificial Intelligence in the field of healthcare: The progress of artificial intelligence is having significant effects in the field of healthcare, providing novel opportunities for the diagnosis, treatment, and care of patients. AI-powered diagnostic technologies have the ability to identify disorders like cancer and cardiovascular diseases in their early stages, which might lead to better patient outcomes (Obermeyer et al., 2019). Algorithms can detect anomalies like tumors in MRI scans or X-rays with high accuracy, often aiding early diagnosis in diseases such as cancer. For instance, Google's DeepMind developed a deep learning model that can diagnose over 50 eye diseases from optical coherence tomography (OCT) scans, assisting doctors in treatment decisions. Machine learning algorithms are used to assess medical pictures, forecast disease outbreaks, and tailor treatment approaches (Heidari 2022). Artificial intelligence (AI) systems have the ability to aid clinicians by detecting patterns in medical data that may not be readily evident to human practitioners (Estava et al., 2017). This enables healthcare workers to dedicate more attention to providing care for their patients.

(e) Robotics and Automation: AI-driven robotics is revolutionizing several sectors by automating monotonous jobs and enhancing accuracy. Robots integrated with artificial intelligence (AI) have the capability to carry out intricate assembly operations, manipulate materials, and conduct precise product inspections in the industrial industry (Brynjolfsson and McAfee 2014). This process of automation has the potential to result in higher levels of production and lower operating expenses. Robots are being used in several fields, including agriculture and logistics, to do duties such as planting, harvesting, and managing inventories (Russell and Norvig 2016). By incorporating artificial intelligence (AI) into robotics, these systems gain the ability to adjust to dynamic situations and carry out tasks independently, resulting in improved efficiency and scalability of operations.

3.2 Ethical Issues raised by modern-day AI advancements

Artificial Intelligence (AI) has evolved significantly over recent decades, transitioning from theoretical concepts to practical applications that impact nearly every aspect of contemporary life as discussed. However, modern-day AI breakthroughs give rise to many ethical dilemmas, some of which are the following:

(a) Confidentiality and Protection of Information:

Artificial intelligence systems often depend on huge quantities of data to operate efficiently. The acquisition and examination of personal data for the purpose of training artificial intelligence models give rise to apprehensions over privacy and monitoring (Zhang et al., 2021). Facial recognition and behavioral monitoring technologies have the potential to violate persons' privacy and result in unlawful surveillance. The ethical dilemma comes in striking a balance between the advantages of data-driven artificial intelligence and the need to safeguard human privacy (Murdoch 2021). Data protection rules, such as the General Data Protection Regulation (GDPR) in the European Union and the California Consumer Privacy Act (CCPA) in the United States, have been implemented to deal with issues around privacy (Bakare et al., 2024). The purpose of these rules is to empower consumers to have more authority over their data and to guarantee that firms that handle personal information follow rigorous requirements. Adhering to these standards is crucial for reducing privacy threats linked to AI. Moreover, companies like Apple and Google use differential privacy to anonymize user data while still enabling AI models to learn from aggregate trends.

(b) Prejudice and Equity

Artificial intelligence systems have the potential to perpetuate or worsen existing biases that are already present in the data used to train them. For instance, partial algorithms used in the process of employment, law enforcement, or lending determinations might result in inequitable treatment of persons based on their ethnicity, gender, or socioeconomic standing (Geiger et al., 2023). The ethical issue at hand is that AI systems have the potential to perpetuate existing structural inequities and give rise to new kinds of discrimination (Dignum 2020). To tackle algorithmic bias, it is necessary to use tactics such as gathering diverse and representative data, ensuring transparency in algorithmic processes, and conducting frequent audits of AI systems (Kerasidou 2021). Efforts to advance equity in AI also include the creation of standards and norms to guarantee that AI technologies do not unfairly penalize any one group. For instance, Algorithmic Impact Assessments (AIAs) are proposed by AI ethics researchers and used by governments like Canada, are tools that assess an AI system's potential impact on equity and discrimination before deployment. These assessments review datasets, monitor model behavior, and analyze risks to reduce prejudiced outcomes.

(c) Responsibility and openness

As artificial intelligence systems gain greater autonomy, the task of assigning responsibility for their judgments gets ever complex. In situations when AI systems produce incorrect or detrimental judgments, it is essential to determine the party accountable for the outcome, whether it be the developers, the users, or the AI system itself (Robinson 2020; Dignum 2020). Creating unambiguous channels of responsibility is crucial to guarantee the responsible use of AI systems. Transparency in AI refers to the ability to comprehend and access the decision-making processes of AI systems (Modi 2023). This include elucidating the functioning of algorithms, the data they use, and the reasoning behind their judgments. Transparency fosters confidence and enables stakeholders to properly assess and question the effects of AI. Many organizations, including Microsoft and Google, have established ethics boards to oversee AI projects. These boards bring together interdisciplinary experts who review and provide guidance on AI initiatives, ensuring they align with ethical principles and accountability standards.

(d) Effects on Employment

Automation driven by developments in AI might result in substantial shifts in the labor market, which may include the displacement of jobs. AI systems are progressively taking over tasks that were historically done by people, which might lead to job displacement and economic upheaval (Russell and Norvig 2016; Suryanarayana 2023). The ethical dilemma is in effectively navigating these changes to provide support for impacted workers and foster economic resilience. To mitigate the effect on employment, it is crucial to allocate resources towards reskilling and education initiatives. These programs aim to provide workers with the necessary skills to adapt to the emerging job prospects brought about by advancements in AI technology. Enacting policies that encourage continuous learning and help workers throughout career changes are crucial for minimizing the adverse consequences of job displacement. Universal Basic Income Proposals or UBI has been proposed as a policy solution to offset job losses caused by AI and automation. While still experimental, UBI pilots have been conducted in countries like Finland and the U.S., providing individuals with a regular income to help transition from jobs replaced by automation.

(e) Ethical Design and Deployment

Integrating ethical issues into the design and development of AI systems is crucial for ethical AI development. This entails ensuring that AI systems are developed with ethical values in consideration, including justice, privacy protection, and responsibility (Robinson 2020; Geiger et al., 2023). Developers and organizations are required to comply with ethical principles and best practices at every stage of the AI lifecycle. It is essential to engage with different stakeholders, such as legislators, industry experts, and the public, to solve ethical problems. Cooperative endeavors may aid in the identification of possible problems, the creation of efficient remedies, and the guarantee that AI technologies conform to society values and ethical norms. For instance, AI Bill of Rights is an example of framework that focused on ethical AI development. Recently proposed by the White House in the United States, this framework lays out guidelines for protecting individuals' rights in AI systems, emphasizing protections against AI-driven discrimination, privacy violations, and opaque practices. It advocates for the fair use of AI and outlines principles that companies should follow when deploying AI technologies.

3.3 AI Ethics: Applied Ethics for Ethical Guidance of AI Advancement and Deployment

With the increasing sophistication and autonomy of AI systems, there are growing worries about their effects on society, privacy, and human rights (Kazim and Koshiyama 2021). This has resulted in the formation of a new discipline called AI ethics, which aims to set standards and norms for the conscientious development and implementation of AI systems. The origins of AI ethics may be traced back to the mid-20th century, coinciding with the emergence of the artificial intelligence field. As AI technologies progressed, ethical questions about their development and use gained more prominence. Alan Turing and other early pioneers posed ethical concerns over the creation of robots capable of displaying intelligent behavior. The emergence of AI ethics as a separate discipline may be credited to the growing intricacy and independence of AI systems, which has sparked discussions on topics like as privacy, prejudice, and responsibility. Significant landmarks in the development of AI ethics include the release of significant publications such as Norbert Wiener's "The Human Use of Human Beings," which established the basis for conversations about the ethical and societal consequences of intelligent computers. The historical backdrop offers useful insights into the development of ethical issues in the field of artificial intelligence research and application.

The ethical issues related to the responsibility and legal liability of AI systems are also brought to the forefront, leading to investigations into who should be held accountable for the activities of autonomous machines (Gabriel 2020). To traverse the complexities of AI development and ensure ethical practices, it

is crucial to create clear ethical frameworks and rules. It is crucial to prioritize the resolution of these significant ethical dilemmas to ensure that AI technologies benefit society and adhere to essential moral standards (Nataliya V. Pashkova et al., 2024).

4. How Cartesian Dualism Can Provide Moral Guidance in Artificial Intelligence:

Descartes' dualism posits a clear distinction between the mental and physical domains. It proposes that the mind (or soul) is an immaterial entity capable of thought, while the body functions within the physical sphere in accordance with natural laws. This distinction has substantial ramifications for comprehending the consciousness of artificial intelligence.

(a) Machine Consciousness: Descartes' perspective suggests that consciousness encompasses an immaterial element, which is absent in existing AI systems that rely on physical hardware and algorithms. Despite its advanced skills, modern AI lacks self-awareness or subjective experience. Descartes' dualism posits that current AI lacks the capability to attain genuine consciousness or self-awareness. This viewpoint strengthens the belief that AI systems should not be regarded as sentient beings with moral entitlements.

The issue of machine experience is brought to attention by Descartes' dualism, which emphasizes the challenge of assigning mental states to machines. Due to its reliance on algorithms and absence of subjective experience, AI does not fulfil the requirements for consciousness as outlined by Cartesian philosophy. This divergence reinforces the contention that ethical deliberations about AI should prioritize its structure and consequences, rather than supposing any manifestation of machine sentience.

(b) Ethics in the Treatment of Artificial Intelligence Entities: While Cartesian Dualism suggests that AI does not possess awareness, it does not exclude the need for ethical issues regarding the creation and use of AI. The ethical handling of AI encompasses several crucial domains:

i. *Effects on Human Well-being*: The implementation of AI systems may have profound ramifications on society, including consequences for employment, privacy, and security. Descartes' philosophy highlights the immaterial essence of the mind, indicating that although AI lacks inherent moral standing, the ethical consequences of its use are of utmost importance. Aligning with ethical concepts that resonate with Cartesian ideals, it is imperative to ensure that AI systems are conceived and deployed in ways that minimize damage and uphold human dignity.

ii. *Design and Accountability:* Descartes' approach emphasizes the need of AI creators to guarantee the ethical use of their products. The primary emphasis should be placed on the design of AI systems, the possible hazards they provide, and the precautions required to alleviate those hazards. The ethical principles and accountability methods for developers demonstrate the moral obligation that comes with inventing technologies that impact human lives.

(c) Concepts of Moral Agency and Responsibility: Descartes' dualism has ramifications for the idea of moral agency in artificial intelligence (AI). Moral agency, as per Cartesian philosophy, requires a conscious intellect that is capable of rational cognition and self-consciousness. Due to the absence of these characteristics, AI systems are incapable of becoming moral agents and should not be subject to the same level of accountability as humans. This differentiation emphasizes the significance of attributing accountability to the human originators and operators of AI systems.

The ethical consequences of AI acts are a direct result of the values and choices made by the persons responsible for designing and operating these systems. Descartes' philosophy asserts that people have the ethical duty, rather than the AI systems themselves. Hence, it is essential to establish comprehensive ethical rules and accountability mechanisms for those engaged in the development and implementation of AI.

(d) Incorporating Cartesian Philosophy with Contemporary AI Ethics: Descartes' theories serve as a philosophical foundation for constructing ethical frameworks that specifically tackle the consequences of AI technology. Cartesian Dualism promotes a clear differentiation between the mental and physical aspects, which directs attention on the effects of AI on humans rather than ascribing awareness or moral significance to the machines.

Considering the continuous advancement of AI technology, it is imperative to engage in continual philosophical contemplation to effectively tackle emerging ethical dilemmas. Cartesian Dualism provides a foundation for exploring these matters, but it is crucial to include other viewpoints and frameworks to comprehensively tackle the intricacies of AI ethics.

5. Conclusion

Exploring Descartes' philosophy allows us to get a profound understanding of the essence of consciousness and morality in artificial intelligence. This investigation may either confirm the Cartesian limits or result in a reinterpretation of consciousness that accounts for the distinctive characteristics of artificial systems. Overcoming these challenges has the potential to close substantial knowledge gaps and provide guidance for ethical frameworks in the development and use of AI. The philosophy of mind developed by René Descartes, namely his notion of Cartesian Dualism, offers useful insights into the ethical problems related to artificial intelligence. Descartes' views elucidate the constraints of contemporary AI systems in terms of awareness and moral agency by differentiating between the mental and physical domains. This framework provides guidance for ethical principles regarding the design, development, and use of AI, with a focus on the responsibility of human creators and users. In order to effectively tackle new difficulties and ensure that AI breakthroughs are in line with human values and social well-being, it will be essential to incorporate Cartesian philosophy with other ethical approaches as AI technology continues to develop.

References

Binns, R. (2018). Fairness in machine learning. *Proceedings of the 2018 ACM Conference on Fairness, Accountability, and Transparency*.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Brynjolfsson, E., & McAfee, A. (2017). *Machine, platform, crowd: Harnessing our digital future*. W. W. Norton & Company.

Cath, C. (2018). Governing artificial intelligence: Ethical, legal, and technical opportunities and threats. *Philosophy & Technology, 31*(4), 521-537. https://doi.org/10.1007/s13347-018-0337-0

Chalmers, D. J. (2010). *The character of consciousness*. Oxford University Press.

Chen, J., & Chien, S. (2019). *Autonomous vehicles: Technology, policy, and practice*. Springer.

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*.

Dennett, D. C. (1991). *Consciousness explained*. (P. Weiner, Illustrator). Little, Brown and Co.

Descartes, R. (1641). *Meditations on first philosophy*. In E. S. Haldane (Trans.), The philosophical works of Descartes (1911 edition). Internet Encyclopedia of Philosophy. Retrieved from http://www.iep.utm.edu

Dreyfus, H. L. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. Artificial Intelligence, 171(6), 1137-1160. https://doi.org/10.1016/j.artint.2007.04.002

Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115-118. https://doi.org/10.1038/nature21056

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Floridi, L. (2019). *The ethics of artificial intelligence*. *Principles, challenges, and opportunities*. Oxford university press.

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014) (pp. 3965-3971). https://doi.org/10.1109/IROS.2014.6942946

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. https://doi.org/10.1038/nature14539

Mulligan, D. K., & Binns, R. (2020). The ethics of artificial intelligence: An overview. In *Handbook of ethics, values, and technological design*.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447-453. https://doi.org/10.1126/science.aax2342

Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Pearson.

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS 2017)*. https://arxiv.org/abs/1706.03762

Blackshaw, B. P. (2023). Artificial consciousness is morally irrelevant. *AJOB Neuroscience, 14*(2), 72-74.

Hildt, E. (2023). The prospects of artificial consciousness: Ethical dimensions and concerns. *AJOB Neuroscience, 14*(2), 58-71.

Cottingham, J. (2008). *Cartesian reflections*. Oxford University Press.

Alanen, L. (1989). Descartes's dualism and the philosophy of mind. *Revue de Métaphysique et de Morale*, 391-413.

Descartes, R. (1954). *Discourse on method*. Harmondsworth, Penguin

Descartes, R. (1951). *Meditations on first philosophy*. Newcomb Livraria Press.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice, and policy. *International Journal of Information Management, 57*, 101994.

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications, and future trends. *Journal of Management Analytics, 6*(1), 1-29.

Aggarwal, C. C. (2018). *Neural networks and deep learning* (Vol. 10, No. 978, p. 3). Cham: Springer.

Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning, and deep learning in advanced robotics: A review. *Cognitive Robotics, 3*, 54-70.

Gupta, N. (2021). Introduction to hardware accelerator systems for artificial intelligence and machine learning. In *Advances in Computers* (Vol. 122, pp. 1-21). Elsevier.

Myers, D., Mohawesh, R., Chellaboina, V. I., Sathvik, A. L., Venkatesh, P., Ho, Y. H., ... & Jararweh, Y. (2024). Foundation and large language models: Fundamentals, challenges, opportunities, and social impacts. *Cluster Computing, 27*(1), 1-26.

Suryanarayana, G. (2023). AI applications of drones. In *Drone Technology: Future Trends and Practical Applications* (p. 153).

Heidari, A., Jafari Navimipour, N., Unal, M., & Toumaj, S. (2022). Machine learning applications for COVID-19 outbreak management. *Neural Computing and Applications, 34*(18), 15313-15348.

Zhang, Y., Wu, M., Tian, G. Y., Zhang, G., & Lu, J. (2021). Ethics and privacy of artificial intelligence: Understandings from bibliometrics. *Knowledge-Based Systems, 222*, 106994.

Murdoch, B. (2021). Privacy and artificial intelligence: Challenges for protecting health information in a new era. *BMC Medical Ethics, 22*, 1-5.

Bakare, S. S., Adeniyi, A. O., Akpuokwe, C. U., & Eneh, N. E. (2024). Data privacy laws and compliance: A comparative review of the EU GDPR and USA regulations. *Computer Science & IT Research Journal, 5*(3), 528-543.

Geiger, R. S., Tandon, U., Gakhokidze, A., Song, L., & Irani, L. (2023). Rethinking artificial intelligence: Algorithmic bias and ethical issues | Making algorithms public: Reimagining auditing from matters of fact to matters of concern. *International Journal of Communication, 18*, 22.

Kerasidou, A. (2021). Ethics of artificial intelligence in global health: Explainability, algorithmic bias, and trust. *Journal of Oral Biology and Craniofacial Research, 11*(4), 612-614.

Modi, T. B. (2023). Artificial intelligence ethics and fairness: A study to address bias and fairness issues in AI systems, and the ethical implications of AI applications. *Revista Review Index Journal of Multidisciplinary, 3*(2), 24-35.

Dignum, V. (2020). Responsibility and artificial intelligence. In *The Oxford Handbook of Ethics of AI* (pp. 4698, 215).

Robinson, S. C. (2020). Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). *Technology in Society, 63*, 101421.