A PHILOSOPHICAL ANALYSIS ON THE TELEOLOGICAL AND DEONTOLOGICAL ETHICAL THEORIES IN ETHICS OF ARTIFICIAL INTELLIGENCE.

By
Uzowulu Onyeka Emmanuel
Department of Philosophy,
Nnamdi Azikiwe University, Awka
Oe.uzowulu@unizik.edu.ng
&
Prof. Charles C. Nweke
Department of Philosophy,
Nnamdi Azikiwe University, Awka
cc.nweke@unizik.edu.ng

Abstract
As artificial intelligence (AI) technologies continue to permeate critical domains such as healthcare, finance, security, and governance, the ethical frameworks guiding their development and deployment have become increasingly vital. This paper examines AI ethics through the lens of two major normative ethical theories: teleological and deontological ethics. Teleological ethics, particularly utilitarianism, evaluates the moral permissibility of AI activities based on their outcomes, prioritizing benefits such as efficiency, scalability, and collective well-being. In contrast, deontological ethics grounds its moral assessment in duties, principles, and the intrinsic rightness or wrongness of actions, irrespective of consequences. The paper explores how each principle addresses key issues in AI ethics, including bias, accountability, privacy, autonomy, and the potential for harm. By analyzing real-world AI applications and ethical dilemmas, this study highlights the strengths and limitations of both perspectives. It argues for a pluralistic approach that balances consequence and sensitive reasoning with principled safeguards, ensuring that AI systems align with both human values and moral responsibility. This ethical synthesis offers a more robust and inclusive foundation for responsible AI governance in an increasingly automated world.

Keywords: Artificial Intelligence, Ethics, Teleology, Deontology, Technology, AI Ethics.

Introduction

In the world today, most people operates at least, one smart device ranging from phones, televisions, computers, automated teller machines (ATM), et cetera. These machines perform required task and respond accordingly simply because they are programs configured and installed by human being and human intelligence. Life was not this easy and simple nevertheless, we have today access to unlimited information. The basic principle controlling machines or modern day devices is strictly the act of an instructed code(s) being stored up as memory in the computer. This breakthrough will not have been possible if not for Alan Turning who in 1935 introduced his abstract "universal computing machine" also known as 'universal Turing Machine' which has metamphorsed into Artificial intelligence (Corry 2017).

A philosophical discussion on artificial intelligence is premised on the assumption that artificial intelligence shares the same vital feature with human beings (intelligence). Precisely, this thinking presupposes that there are certain features of intelligence stimulated by machines with a potential of advancing in the nearest future. Judging from previous researches where it is accepted that only humans were capable of intelligence and reasoning, the introduction of AI

has raised some philosophical, ethical, existential and ontological issues about the nature of the mind. Initially, most thinkers thought that machines cannot be like humans because they (machine) cannot think. Today, it is evident that machines can think by simulating some characteristics of intelligence through different means. The debate today takes a U-turn from whether machines can think to whether machines can possess mind? Scientists and researchers have succeeded in creating artificial intelligence which possesses a super-intelligent mind such as self-driving cars, Alexia and Siri.

There is an existential fear and real disgust by many researchers who view that machines possessing super-intelligent abilities will in no distant time mature into a threat to mankind. George Dvorsky (2008), pointed out that a time in the future will come when artificial intelligence will rise above humanity. He asked; when that happens, what are we to do? With the many surroundings on AI, contemporary researchers and scientists around the world are advancing into deeper research and investigations on issues of AI technology. The future is an unknown reality and for some, it holds expectations of wonderful things. With recent advancement of artificial intelligence, the hope of immortality is on the rise especially with the merging of humans to machines lately. Diamandis and Kotler (2020) recognizes that AI will be the key to usher in the era of abundance where there would be enough food, water, shelter and rest. However, skeptics like Brynjolfsson, Rock, and Syverson (2017) express their worry about the state of employment and the consequences of AI and Robotics.

Barrat (2023) in discussing the dangers of AI makes a convincing case when he writes; "I was convinced that the knowledgeable experts who did not question AI's safety at all were suffering from delusions." He recognizes that there will be a real breakthrough in AI leading to artificial super intelligence (ASI) which according to him is an intelligence much greater than human intelligence. The ASI is superior to the (AGI) Artificial General Intelligence which is intelligence at the human level. The flow of progression in Barat's analysis shows that AI (phones, internet, appliances) has progressed to AGI and then to ASI.

Artificial intelligence's potential effects on human existence remains a topic of intense discussion. While some academics and researchers continue to voice their conviction and optimism that artificial intelligence, with its immense potential, can improve the economy and lessen human stress, others are concerned about its unfavorable effects, which could go beyond the point of no return and endanger humankind.

Philosophy's inherent focus is problem-solving, or formulating enduring answers to the existential problems that face humanity everywhere and at any time. The introduction of artificial intelligence (AI) into many areas of society has brought forth amazing breakthroughs and revolutionary possibilities. This quick development of AI technologies brings up serious ethical issues and complex problems. The rapid integration of Artificial Intelligence (AI) into diverse sectors ranging from healthcare and finance to education, security, and governance, has triggered a profound need for rigorous ethical scrutiny. As AI systems increasingly make or influence decisions that affect human lives, questions about fairness, accountability, autonomy, and the potential for harm have moved to the forefront of scholarly and policy debates. In addressing these concerns, normative ethical theories provide vital frameworks for assessing the morality of AI design, implementation, and impact. Among these, teleological (consequence-based) and deontological (duty-based) ethics remain two of the most influential paradigms in moral philosophy.

Teleological ethics, particularly in its utilitarian form, evaluates actions based on their outcomes, focusing on maximizing overall well-being or minimizing harm. This framework

has proven appealing in AI contexts where efficiency, scalability, and social benefit are often primary goals. Conversely, deontological ethics judges the morality of actions based on adherence to universal duties, principles, or rights regardless of the consequences. This theory emphasizes respect for individual autonomy, justice, and the intrinsic value of persons, making it particularly relevant to concerns around algorithmic bias, privacy, and the erosion of human dignity.

This paper critically examines how these two ethical theories address key moral challenges posed by AI systems. Specifically, it analyzes the role of teleological and deontological reasoning in relation to bias**,** accountability**,** privacy**,** autonomy**,** and the potential for harm. By exploring the strengths and limitations of each approach, the paper argues for a more integrated and context-sensitive ethical framework that balances principled constraints with outcome-based reasoning. Such a hybrid approach is essential for guiding the responsible development and governance of AI in a manner that aligns with both societal values and fundamental moral obligations.

WHAT IS ARTIFICIAL INTELLIGENCE?

Artificial Intelligence (AI) is one of the most groundbreaking and fast-developing technologies of the 21st century (Tamuno Miegbam and Bariledum 2022). It involves creating machines that can mimic human intelligence by thinking, learning, and making decisions. AI systems are built to handle tasks that usually require human cognition, such as image recognition, speech processing, decision-making, and language translation. From autonomous vehicles to digital assistants like Siri and Alexa, AI is steadily integrating into everyday life, transforming industries, economies, and society as a whole.

One of the fascinating aspects of the field of artificial intelligence (AI) is that the precise nature of its subject matter turns out to be surprisingly difficult to define. The problem involves securing an adequate grasp of the nature of the artificial intelligence. First, the question, 'What is supposed to be "artificial" about artificial intelligence?' with no doubt, has to do with its origins and mode of creation in arising as a product of human contrivance and ingenuity rather than as a result of natural (especially biological or evolutionary) influence. Things that are *artificially intelligent*, in other words, differ from those that are *naturally intelligent* as artifacts that possess special properties ordinarily possessed by non-artifacts. So these are things that have a certain property (intelligence) as a result of a certain process (because they were created, designed, or manufactured in this way).

These artifacts according to James Fetzer (2001), are of a certain special kind in the sense that they are commonly thought of as being *machines.* If machines are simply things that are capable of performing work, then, since human beings are capable of performing work, human beings turn out to be machines, too. If anyone wants to know whether or not there are such things as intelligent machines, therefore, the answer seems obvious so long as human beings are (at least, some of the time) *intelligent.* A more difficult question arises from distinguishing between animate and inanimate machines. Human beings, after all, may result from human contrivance and ingenuity, but they are clearly biological in their origin. The issue thus becomes whether or not *inanimate* machines, as opposed to human beings, are capable of possessing a certain special property that human beings are supposed to display and if not always, at least on certain special occasions. The problem revolves about the identification or the definition of what is meant to be "intelligent" about artificial intelligence. The dictionary may seem to be an appropriate place to begin.

Webster's New World Dictionary (1988), defines "intelligence" as the ability to learn or understand from experience; ability to acquire and retain knowledge; mental ability; b) the ability to respond quickly and successfully to a new situation; use of the faculty of reason in solving problems, directing conduct, etc., effectively; C) in *psychology*, measured success in using these abilities to perform certain tasks". If the concept of intelligence is to be applicable (at least, in principle) to inanimate machines, however, it must not be the case that inanimate machines could not possibly be intelligent as "a matter of definition".

Definitions are often viewed as arbitrary or subject to agreement, as if their resolution depends solely on who has authority. However, the meanings we assign to words play a crucial role in how we address problems and seek solutions. In this context, defining "intelligence" is essentially equivalent to developing a theory about its nature as a real-world phenomenon. A definition that either excludes humans from the category of intelligent beings or includes objects like tables and chairs would clearly be inadequate. By examining clear examples where intelligence is either undeniably present or absent, we can refine our understanding of its essential characteristics, even if full consensus remains elusive, a process akin to prototype theory. Even if we accept that humans are intelligent, the challenge lies in determining which aspects of human existence constitute intelligence. Since humans experience emotions such as anger, jealousy, and rage, one might question whether machines displaying these emotions should be considered intelligent. The issue is not merely that people might mistakenly associate these emotions with intelligence, rather, unless we already know they are unrelated, we cannot exclude them from consideration. If humans are our primary example of intelligence and frequently display such emotions, then what criteria allow us to determine that these emotions are not integral to intelligence? In fact, there may be specific conditions under which these emotions could be considered relevant to intelligence.

There are multiple ways to define artificial intelligence (AI), but a useful starting point is to examine the kinds of problems AI technology is designed to solve. In this sense, AI can be described as the use of technology to automate tasks that typically require human intelligence. This definition highlights AI's focus on replicating human-like cognitive abilities to perform specific tasks. Bernard Marr (2024) introduced several examples that illustrate this concept. Accordingly, AI has been successfully applied to complex activities such as playing chess, translating languages, and driving vehicles. But what distinguishes AI-driven tasks from general automation? The key distinction lies in the fact that when humans perform these tasks, they rely on advanced cognitive processes associated with intelligence. For example, playing chess requires reasoning, strategy, planning, and decision-making. Language translation involves interpreting symbols, understanding context, and processing meaning. Similarly, driving a car engages multiple cognitive functions, including vision, spatial awareness, situational assessment, motor control, and judgment. These examples demonstrate how AI aims to replicate the intellectual processes that humans naturally employ in these activities.

Many people mistakenly perceive AI as a form of machine intelligence that thinks like a human. A common misconception is that current AI systems achieve their results through advanced synthetic cognition that equals or surpasses human reasoning. In reality, today's AI systems are not genuinely intelligent in the way humans are. Instead, AI can generate intelligent-looking outcomes without actually possessing intelligence. These systems primarily rely on heuristics, identifying patterns in data and applying predefined rules, knowledge, and information programmed by humans into computer-readable formats. By leveraging these computational techniques, AI can often perform complex tasks that typically require human cognition.

However, it is important to note that while AI may produce impressive results, the underlying mechanisms it uses are fundamentally different from human thought processes.

The term *artificial intelligence* was introduced by John McCarthy in 1956 when he organized the first academic conference on the subject (Bhutani, Aditi 2023). However, the exploration of whether machines can truly think began much earlier. In his influential work *As We May Think*, Vannevar Bush (1945) envisioned a system designed to enhance human knowledge and comprehension. A few years later, Alan Turing expanded on this idea, proposing that machines could simulate human behavior and perform intelligent tasks, such as playing chess.

While there is no doubt that computers can process logical operations, whether they can actually *think* remains an open question. The definition of "thinking" is crucial, as many have strongly debated whether it is even possible for machines to possess this ability. One notable challenge is the *Chinese Room* argument, introduced by John Searle in 1980. This thought experiment imagines a person locked in a room, receiving notes written in Chinese. By following an extensive set of rules and lookup tables, they can generate appropriate responses without truly understanding the language. The argument suggests that, similarly, computers relying solely on predefined rules and data retrieval may never achieve genuine understanding. Although this argument has been widely challenged by researchers, it continues to influence skepticism regarding AI's capabilities, particularly in critical applications that require expert decision-making.

Over the past sixty years, significant progress in AI has primarily come from improvements in search algorithms, machine learning techniques, and the integration of statistical analysis into various fields. However, many of these advancements go unnoticed by the general public. Instead of the futuristic AI imagined in science fiction such as talking machines piloting spaceships, AI is more commonly used in subtle applications like analyzing purchase histories and influencing marketing strategies. In the opinion of Nowak, Andrzej et.al (2018), states that many consider to be *true AI* has not advanced as rapidly as expected. A recurring theme in the field has been the tendency to underestimate the difficulty of fundamental challenges. For decades, major AI breakthroughs have been promised to arrive "within ten years," yet these predictions have repeatedly fallen short. Additionally, there is a persistent tendency to redefine what qualifies as "intelligence" each time machines achieve a new milestone, an issue known as the *AI Effect*. This shifting goalpost contributed to the decline of AI research in the 1980s. From the AI effect which is a common pattern across all AI types is that once AI becomes widely adopted, it often stops being recognized as AI. This majorly occurs when people dismiss AI's capabilities, arguing that they do not represent true intelligence. As British science fiction writer Arthur C. Clarke famously stated, "any sufficiently advanced technology is indistinguishable from magic." By implication, once the underlying technology is understood, the sense of magic fades.

In 1950, British mathematician Alan Turing (2009) published a groundbreaking paper titled *Computing Machinery and Intelligence*, which laid the foundation for what would later be known as artificial intelligence, years before John McCarthy officially coined the term. In the paper, Turing posed a fundamental question: *Can machines think?* He then introduced a method for assessing machine intelligence, which later became known as the Turing Test. Originally referred to as the *Imitation Game*, the test was designed as a straightforward way to determine whether a machine could be considered capable of thought. Turing's approach was pragmatic that is, if a computer could successfully mimic human intelligence to the point of being indistinguishable from a person, it could be said to demonstrate thinking capabilities.

The Turing Test remains a fundamental long-term objective in AI research. Majorly, can we develop a machine capable of mimicking human behavior so convincingly that even a skeptical judge cannot distinguish between human and AI? Since its introduction, the test has mirrored the trajectory of AI research as a whole. Initially, it seemed like a challenging but achievable goal, dependent primarily on advancements in hardware. However, as research progressed, it became clear that the problem was far more complex than originally anticipated, leading to a slowdown in progress and skepticism about whether it will ever be fully achieved. Despite significant technological advancements over the decades, the Turing Test continues to serve as a benchmark, reminding researchers just how far AI still has to go before reaching true human-like intelligence.

Expectations in AI often surpass reality. Despite decades of research, no AI system has successfully passed the Turing Test, expert systems have expanded but remain less widespread than human experts, and while AI can outperform humans in certain games, it still struggles with open-ended gameplay (Moor 2012). This raises a critical question: Have we simply failed to allocate enough resources to fundamental AI research, as seen in past AI winters? Or is AI's complexity something we have yet to fully grasp, leading researchers to focus on specialized tasks such as mastering chess rather than truly understanding intelligence itself?

AI can be categorized in different ways based on the type of intelligence it exhibits or its stage of development. It can be classified into *analytical, human-inspired,* and *humanized AI* depending on whether it demonstrates cognitive, emotional, or social intelligence. Alternatively, AI can also be divided into *Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI),* and *Artificial Super Intelligence (ASI)* based on its level of advancement (Fahad et al. 2024). Where AI – Artificial Intelligence is at the level of phones, internet; AGI – Artificial General Intelligence is at the level of man (humans); ASI - Artificial Super Intelligence is much greater than human intelligence

WHAT IS AI ETHICS?

Artificial Intelligence (AI) is rapidly transforming societies by influencing decision-making in healthcare, finance, education, business, security, and even politics. As machines acquire greater autonomy and decision-making capacities, questions about the moral implications of their design and use have become unavoidable. AI ethics emerges as the branch of applied ethics and philosophy that examines these questions, offering frameworks to guide the responsible development and deployment of artificial intelligence. AI ethics refers to the study and application of moral principles to the design, development, and use of artificial intelligence systems. It interrogates how values such as fairness, justice, accountability, privacy, autonomy, and human well-being should be incorporated into AI technologies. Unlike traditional computing ethics, which largely focused on human actions when using technology, AI ethics emphasizes the behavior of autonomous or semi-autonomous systems themselves and the responsibilities of those who create and deploy them.

AI ethics should be approached through a multidisciplinary lens that integrates rigorous, context-specific inquiry within the broader ethical discourse on artificial intelligence. This involves not only fostering substantive academic and policy engagement with the ethical dimensions of AI, but also treating AI ethics as a dynamic and evolving system grounded in virtue ethics. Such a framework encourages all stakeholders, developers, policymakers, corporations, and end-users to cultivate moral responsibility and to hold one another accountable for the societal impacts of AI systems. The research ultimately argues for

conceptualizing AI ethics as situated along a continuum of intersecting and interdependent interests across academia, civil society, public governance, and the private sector ranging from early-stage tech start-ups to multinational corporations. Thus, this is a call that emphasizes that actionable and transparent ethical practices are not only necessary for legitimacy but can also constitute a form of institutional and societal value.

Amid the widespread inefficacy of many recent AI policy initiatives including intergovernmental agreements, national strategies, professional guidelines, and private-sector standards, there is a growing recognition within the AI research and policy communities of the urgent need to operationalize AI ethics in concrete, enforceable ways. Luciano Floridi (2019) in his essay *Translating Principles into Practices of Digital Ethics*, outlines key dangers that arise when AI ethics remains abstract, idealistic, or disconnected from the realities of technical implementation. Floridi (2019), identifies five problematic trends that threaten the credibility and efficacy of ethical governance in AI; Ethics shopping which is the selective adoption of ethical principles from an overwhelming number of competing frameworks in the absence of unified and publicly endorsed ethical standards. Ethics bluewashing which is the superficial or performative commitments to AI ethics without establishing verifiable, evidence-based mechanisms of accountability and transparency, particularly among influential actors such as corporate boards and government officials. Ethics lobbying is the prioritization of self-regulation over binding legal and ethical oversight. Ethics dumping involves the outsourcing of ethically questionable research practices to jurisdictions with weaker regulatory environments, undermining efforts to build a global culture of ethical research and consumption. Lastly, Ethics shirking which involves the neglect of ethical responsibilities due to perceived low incentives, in contrast to the need for well-defined accountability structures.

Brent Mittelstadt (2019), in his article *Principles Alone Cannot Guarantee Ethical AI*, argues that the substantive work of AI ethics begins only now, in the wake of a proliferation of ethical principles and policy frameworks. He contends that the real challenge lies in the translation of these abstract principles into concrete practices, which will, in turn, illuminate the genuine ethical complexities posed by AI technologies. Moreover, AI ethics is grounded on a foundation of mutual respect across a wide range of disciplines including philosophy, computer science, anthropology, statistics, political science, law, and mathematics. If this same interdisciplinary respect informs how the technical community engages with AI ethics, the likelihood of translating ethical principles into meaningful action increases significantly. Such an approach would foster a shared understanding of the societal implications of AI technologies and encourage a collective commitment to ethical responsibility.

In the realm of AI ethics, Vincent C. Müller (2025), writing for the *Philosophy of AI: A structured overview,* delineates two principal domains: one concerns the ethical implications of AI systems as objects, tools created and utilized by humans while the other, addresses the ethical treatment of AI systems as potential subjects in their own right. One can easily propose, however, that a third dimension deserves attention: the conceptualization of AI as part of a technological imaginary and the ethical consequences that arise from this symbolic framing. Emerging scholarship has begun to explore this terrain. A good example is how Cave and Dihal (2021) highlight the representations of AI as overwhelmingly coded as white, both in visual depiction and cultural associations, revealing underlying racialized narratives embedded in the public imagination of AI.

Similarly, Thilo Hagendorff (2020), in *The Ethics of AI Ethics: An Evaluation of Ethical Guidelines*, echoes this concern by highlighting several critical limitations within current AI

ethics discourse. He points out that AI ethics often lacks the structural mechanisms necessary to uphold and enforce its normative claims. One major issue is the perception of ethical guidelines as external impositions on the technical community, which contributes to their limited influence. Furthermore, Hagendorff (2020) identifies the absence of distributed responsibility and the widespread lack of awareness among software developers regarding the broader, long-term societal impacts of their work. This disconnect, he argues, undermines the developers' sense of accountability and diminishes the ethical salience of their roles.

What, then, is meant by the ethics of AI? Broadly speaking, it can be divided into two overlapping areas: one concerned with the ethical dimensions of machines and computational systems themselves, and the other with the societal implications of AI technologies. Sometimes, AI ethics is equated with computer ethics, as reflected in Floridi, Luciano's foundational definition of computer ethics as the "analysis of the nature and social impact of computer technology and the corresponding formulation and justification of policies for the ethical use of technology (Palm and Hansson 2006)." Such a definition encompasses AI ethics within the broader discipline of technology ethics. To make AI ethics more accessible and practically relevant especially for the technical community and civil society this research proposes understanding AI ethics as a policymaking domain shaped by concerns raised across its various subfields. The diversity of definitions and approaches to AI ethics is partly due to the field's evolution into several overlapping areas, including machine ethics, algorithm or data ethics, robot ethics, information ethics, and neuro-ethics.

Machine ethics is primarily concerned with the moral dimensions of designing artificial systems capable of ethical decision-making and with analyzing artificial morality from a socio-ethical perspective. David J Gunkel (2012), distinguishes between machine ethics and computer ethics: the latter focuses on human actions mediated by digital technologies, while the former considers machines themselves as potential moral agents. In this context, machine ethics seeks to address "ethics for machines", treating machines as subjects, not merely tools for human use. Robot ethics and computer ethics, by contrast, focus more on human responsibilities when designing and interacting with such systems. Samuel T Segun (2021), suggests further refining machine ethics into two domains: computational ethics (addressing technical feasibility) and machine ethics proper (addressing moral agency).

Robot ethics, or the ethics of social robots, examines how robots influence society through human-robot interactions (HRI), the tendency to anthropomorphize robots, and ethical concerns around robot rights and the objectification of humans. This subfield can be further categorized by analyzing AI-human, AI-AI, and AI-society interactions. Closely related is robo-ethics, which considers the ethical responsibilities of humans who design, construct, and deploy AI agents. Emma Ruttkamp-Bloem (2020) proposes a comprehensive model of 'robot ethics' that includes: (1) the ethical systems embedded in robots (linking with machine ethics), (2) the ethics of those who create and use robots (aligned with computer ethics and robo-ethics), and (3) the ethics of how humans relate to robots (which overlaps with the ethics of social robots). She argues that the most effective framework treats robots as socio-technical systems, addressing all three dimensions together.

Many scholars in AI ethics tend to treat the rise of artificial intelligence as inevitable, a natural progression in technological development. They often operate under the assumption that AI is inherently superior to previous technologies, framing the ethical task primarily as making AI "ethical" or "trustworthy." However, this perspective is flawed. The ultimate objective should not be ethical AI in itself, but rather the creation of a just, peaceful, and sustainable society.

Ethical norms should emerge from this broader social vision. Therefore, AI ethics must critically question where and whether AI with its reliance on prediction and classification, should be used at all within a truly ethical society. A significant shortcoming of current AI ethics is its tendency to accept the increasing integration of AI into various sectors without challenging the underlying dysfunctional systems it reinforces. Together, these critiques emphasize that without robust enforcement structures, internal cultural integration within the technical community, and a more deeply embedded understanding of the societal stakes involved, current AI ethics risks remaining aspirational rather than actionable.

DEONTOLOGICAL ETHICAL PERSPECTIVE ON AI ETHICS

The philosopher Immanuel Kant introduced the idea of deontological ethics, which is often referred to as Kantian deontology (Benlahcene et al. 2018). As a committed Christian, Kant based his ethical theory on universal moral duties. He also believed that every human being possesses intrinsic worth, and thus, concepts such as autonomy, dignity, and respect for individuals must be upheld. Later, Ross expanded on Kant's framework by introducing a more flexible version of deontology that includes multiple moral duties, such as avoiding harm and keeping promises (Browning 2015). Unlike utilitarian ethics, which focuses on the outcomes of actions, deontology centers on moral obligations prohibiting actions like causing harm, even if doing so could result in beneficial outcomes. This means that choices grounded in deontological principles may be ethically correct for an individual even if they don't produce the best results for society.

While utilitarianism assesses the morality of an action by its potential to produce the greatest good for the majority, deontology judges an action based on its inherent nature, not its consequences (Wesseh 2025). For instance, Major General McClintock and Brigadier General Ford engineered and retained a virus as a biological weapon in hopes of winning a war (a desired outcome). However, creating and storing the virus was intrinsically unethical and led to widespread death, thus violating both utilitarian and deontological ethics.

Deontological ethics places a strong emphasis on the inherent value of human life and upholds principles such as respect for autonomy, beneficence, non-maleficence, and justice. These values are especially relevant in moral decisions where deontological ethics provides a framework for interpreting the four foundational ethical principles: respecting autonomy, avoiding harm, ensuring justice (as expressions of human dignity), and promoting well-being (as an expression of beneficence).

Kant argued that individuals must act according to moral duty, even if it involves personal risk. The morality of an action, in his view, depends not on its outcomes but on the intention behind it. Therefore, in deontological ethics, moral decision-making is guided by one's sense of duty and obligation rather than by the potential consequences of an action (Johnson and Cureton 2004). Deontological ethics is rooted in universal moral duties that can be understood through reason, allowing individuals to instinctively recognize immoral situations. For example, it would be morally wrong to sacrifice infected individuals to protect the uninfected. From a deontological perspective, healthcare professionals have a duty to support those who are already infected, treating them with dignity, avoiding harm, and showing respect and compassion while fulfilling their ethical and medical responsibilities.

TELEOLOGICAL ETHICAL PERSPECTIVE ON AI ETHICS

Teleological ethical theories are grounded in the idea that the morality of an action depends on evaluating its outcomes. The term "teleology" comes from the Greek word *telos*, meaning "end" or "goal" (McKitterick 2022). According to this perspective, actions are not inherently right or wrong; rather, it is their consequences for those affected that determine their moral value. Utilitarianism, a well-known teleological theory, is commonly used in economics to explain rational decision-making (discussed in more detail in later sections). Jeremy Bentham, an 18th-century philosopher, and John Stuart Mill, a 19th-century thinker, are key figures in the development of Utilitarianism, with Mill credited for coining the term. This theory holds that human morality is driven by the aim to maximize satisfaction and efficiency. Therefore, an action is considered morally right if it leads to greater happiness and wrong if it results in the opposite.
As Bentham famously stated; "The greatest happiness of the greatest number is the foundation of morality and legislation" (Veenhoven 2014). A compelling question raised within Utilitarianism is: "What exactly is happiness?" Initially, happiness was understood simply as physical or sensory pleasure. This interpretation led to sharp criticism of Bentham, who, under such a view, seemed to suggest that it was preferable to be "a satisfied pig than a discontented Socrates" (Boatright 2010). John Stuart Mill expanded on Bentham's ideas by introducing the concept of a hierarchy of pleasures. He argued that it is "better to be a dissatisfied human than a satisfied pig". As a result, Mill's version of Utilitarianism extended beyond mere material or bodily pleasures. The theory also assumes that happiness can be quantified, meaning both the amount and the quality of happiness can be assessed and compared.

TELEOLOGICAL AND DEONTOLOGICAL ETHICAL THEORIES IN AI ETHICS: A PHILOSOPHICAL ANALYSIS.

The deontological ethical perspective emphasizes the inherent rightness or wrongness of an action, regardless of its consequences. One example of this is the Doctrine of Double Effect, which suggests that an action may be ethically acceptable if its positive outcomes outweigh its negative ones. Applied to AI, this principle implies that the use of AI technologies like ChatGPT, Meta AI, SIRI, et cetera, can be ethically justified if the benefits such as increased efficiency and support in completing tasks are greater than the potential harms, such as encouraging laziness. Ross's theory of *prima facie* duties can be related to AI role. Accordingly, individuals have a moral obligation to uphold the integrity and authenticity of their work. For example, using AI tools to reproduce or submit work that is not their own may conflict with the duty to maintain academic honesty. While AI tools can assist with learning and writing, its use must remain within ethical boundaries to avoid breaching these responsibilities (Hosseini, Resnik, and Holmes 2023).

A rights-based ethical framework becomes relevant. Individuals are entitled to freedom and equal access to knowledge, but these rights must not come at the expense of others. Using AI tools in ways that violate the intellectual property rights of others or undermine fair competition may be considered unethical. Hence, one must use AI tools in a way that respects the rights of peers and original creators.

On one hand, teleological ethics which is derived from the Greek *telos* (end) and *logos* (reasoning), focuses on achieving morally desirable goals rather than adhering to fixed rules. Unlike deontology on the other hand, which evaluates actions based on principles regardless of outcomes, teleology assesses actions based on their alignment with meaningful ends.

Though similar to consequentialism, teleology centers more on the intended goal than on the broader consequences. Utilitarianism, for example, blends both views. In the case of AI tools, a teleological perspective might support its use if it contributes positively to an individual development. Thus, if AI tools enable a person to understand concepts more or successfully complete their tasks, its use can be seen as ethically justifiable within this framework.

The uncertainties inherent in intuition-based ethics can be largely mitigated through a systematic application of deontological ethics to machines. While "deontological ethics" literally means duty-based ethics, in practice, it is commonly understood as rule-based ethics where moral obligations are expressed as specific rules of conduct. Although this tradition is closely associated with Immanuel Kant, its application is not restricted to his historical framework. The fundamental logic underpinning Kantian ethics can be extended to generate clear, principle-based rules. A central insight here is that all deliberate actions are grounded in reasons. While behavior may be driven by physical or biological causes, a "dual standpoint" approach helps distinguish between mere behavior and genuine action, action being understood as something that can be explained by an agent's reasons. Therefore, any given action can be interpreted both as the result of physical causation and as the outcome of a reasoning process.

De Colle and Werhane (2008), in *Moral motivation across ethical theories*, believes that since actions are grounded in reasons, they typically follow a conditional structure, "if these reasons apply, then perform this action." These kinds of conditional rules are well-suited for encoding in machines. Furthermore, by requiring that the reasons behind actions meet certain standards of logical consistency, a core concern of deontological ethics then one can derive valid ethical rules. Importantly, this does not assume that machines are autonomous moral agents in the full philosophical sense. Rather, machines are considered autonomous only in a limited, operational sense they follow internal rules without continuous human oversight. Therefore, the ethical responsibility lies with the human programmer, who must ensure that the rules embedded within the machine are ethically justified. The machine, therefore, acts in accordance with rules that the human designer is morally accountable for introducing.

Using the table below, the researchers provide a critical analysis that explores how each theory engages with key issues in AI ethics ranging from bias, accountability, privacy, autonomy, and potential harm.

Ethical Issues	Teleological Perspective	Deontological Perspective
Bias	Teleological ethics, especially utilitarianism, evaluates AI bias in terms of its *impact on overall well-being*. If biased algorithms disproportionately harm marginalized communities such as in facial recognition or hiring systems, utilitarianism would deem them unethical because they reduce collective happiness and social utility. The utilitarian goal is to *maximize fairness and accuracy* across all users to ensure the greatest benefit to society. However, this approach may inadvertently justify some degree of bias if it results in a greater good for the majority, raising concerns about sacrificing minority rights.	Deontological ethics rejects bias categorically because it violates moral duties of fairness, equality, and respect for persons. For Kantian ethics, treating individuals as mere means to an end (e.g., optimizing a system at their expense) is morally impermissible, regardless of overall benefits. A biased algorithm infringes on the duty of impartiality and the principle of justice, making it unethical in deontological terms *even if it benefits the majority*. Thus, this offers a stronger safeguard against discriminatory AI practices.
Accountability	From a teleological viewpoint, accountability is instrumentally valuable insofar as it leads to better outcomes, such as trust, safety, or system improvements. If delegating decision-making to an AI improves efficiency without significant negative consequences, it may be deemed acceptable. However, utilitarianism struggles with diffused responsibility, particularly when harm occurs due to complex, opaque algorithmic decisions involving multiple actors (developers, users, data providers). This can blur lines of moral responsibility.	Deontology holds that accountability is a moral duty, regardless of outcomes. Each agent involved in AI creation or deployment has a responsibility to act ethically and take ownership of their decisions. Deontology insists on clear lines of moral agency, rejecting the idea that responsibility can be entirely delegated to machines. Deontology thus supports strong ethical design obligations, including explainability, human oversight, and moral traceability.
Privacy	A teleological ethicist would argue that privacy can be overridden if doing so produces a net benefit such as improved healthcare diagnostics through data aggregation or enhanced national security through surveillance. The cost-benefit calculation determines the ethical permissibility of data collection or monitoring. However, this can lead to ethical compromises, especially when the benefits to the majority	For deontology, privacy is an inviolable right grounded in the duty to respect individuals as autonomous moral agents. Unauthorized data collection, surveillance, or profiling constitute a moral violation, regardless of the good it might achieve. Deontology therefore demands consent, transparency, and respect for informed self-determination. It resists utilitarian justifications for invasive

	overshadow the privacy rights of individuals.	technologies, making it particularly robust in defending individual privacy in the age of AI.
Autonomy	Utilitarian ethics values autonomy instrumentally as a means to achieving well-being. AI systems that assist or nudge users (e.g., recommender systems, behavior modification tools) may be justified if they enhance satisfaction or reduce harm. However, this opens the door to manipulation or paternalism, where autonomy is sacrificed for presumed utility, leading to soft coercion (e.g., algorithmic "nudging" without informed consent).	Autonomy is central to deontological ethics. AI systems must be designed in ways that respect human agency, informed choice, and freedom from manipulation. Any system that undermines these values, no matter how beneficial, would be considered unethical. Deontological ethics promotes explainable AI and user-centered design as moral imperatives, ensuring that humans remain the authors of their own decisions.
Potential for Harm	Teleological ethics directly weighs harms against benefits. If AI causes harm but ultimately prevents greater harm or produces substantial benefit (e.g., in military defense or medical triage), it may be ethically acceptable. The principle of "the greatest good for the greatest number" dominates, yet this can justify collateral damage, as long as the total utility is positive. This approach is pragmatic but ethically risky in high-stakes AI applications.	Deontological ethics prohibits intentional harm, regardless of benefits. Even if an AI system saves lives, it is unethical if it causes intentional or foreseeable harm in the process. This is grounded in the duty of non-maleficence which is, do no harm. Thus, AI systems must be designed with strict ethical constraints, and harms cannot be justified by outcomes. This ensures high moral integrity but may limit innovation or practical applications in urgent contexts.

From the above, artificial intelligence (AI) introduces complex ethical dilemmas, challenging traditional moral frameworks to provide guidance on how AI systems should be designed, deployed, and governed. The two foundational ethical theories; teleological ethics (consequentialism) and deontological ethics (duty-based ethics), so far have offered distinct approaches to evaluating and resolving these challenges.

CONCLUSION

Most of the discourse is grounded in two dominant ethical theories: deontology and consequentialism. Notably, there is a significant lack of empirical research in the field, with much of the focus concentrated on theorizing or technical design, rather than on practical validation. This overreliance on deontological and consequentialist approaches risks embedding a narrow logic. There remains a critical need for more engagements to evaluate how various ethical theories perform in practice and to identify those that are most suitable to guide AI ethics especially in real-world decision-making scenarios. The current ethical framing also limits our understanding of the broader relational and existential dimensions of human-

machine interaction. Questions such as “What constitutes a good relationship between humans and machines?” or “What new forms of moral agency might arise through our interactions with intelligent systems?” are largely neglected.

Both teleological and deontological theories offer valuable yet incomplete insights for AI ethics. Teleology provides flexibility and context-sensitive reasoning, but may risk justifying harm or injustice. Deontology offers principled protection of rights and duties, but can be overly rigid or inapplicable in real-world dilemmas. A hybrid approach, combining outcome sensitivity with non-negotiable moral duties, may offer a more comprehensive ethical foundation. As AI technologies become more autonomous and embedded in human life, ethical frameworks must evolve to reflect both moral reasoning and practical responsibility.

REFERENCES

Barrat, James. 2023. *Our final invention: Artificial intelligence and the end of the human era*. Hachette UK.

Benlahcene, Abderrahmane, Ruslan Bin Zainuddin, Nur Syakiran, and Akmal Bt Ismail. 2018. "A narrative review of ethics theories: Teleological & deontological ethics." *Journal of Humanities and Social Science (IOSR-JHSS)* 23 (1): 31-32.

Boatright, John R. 2010. *Finance ethics: Critical issues in theory and practice*. John Wiley & Sons.

Browning, Nicholas. 2015. "The ethics of two-way symmetry and the dilemmas of dialogic Kantianism." *Journal of Media Ethics* 30 (1): 3-18.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2017. *Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics.* National Bureau of Economic Research.

Bhutani, Aditi, and A Snaria. 2023. "The past, present and future in artificial intelligence." *GLMS J Manag Rev Transform* 2: 132-140.

Bush, Vannevar. (1945): "As we may think." *The atlantic monthly* 176, no. 1, 101-108.

Cave, Stephen, and Kanta Dihal. 2021. "Race and AI: The diversity dilemma." *Philosophy & Technology* 34 (4): 1775-1779.

Clarke, Arthur Charles, and Stephen Banker. 1987. *Arthur C. Clarke*. Octopus/Heinemann.

Corry, Leo. 2017. "Turing's pre-war analog computers: The fatherhood of the modern computer revisited." *Communications of the ACM* 60 (8): 50-58.

De Colle, Simone, and Patricia H Werhane. 2008. "Moral motivation across ethical theories: What can we learn for designing corporate ethics programs?" *Journal of Business Ethics* 81 (4): 751-764.

Diamandis, Peter H, and Steven Kotler. 2020. *The future is faster than you think: How converging technologies are transforming business, industries, and our lives*. Simon & Schuster.

Dvorsky, George. 2008. "Better Living through Transhumanism." *Journal of Evolution & Technology* 19 (1).

Fahad, Muhammad, Tayyaba Basri, Muhammad Ameer Hamza, et. al., 2024. "The benefits and risks of artificial general intelligence (AGI)." In *Artificial general intelligence (AGI) security: Smart applications and sustainable technologies*, 27-52. Springer.

Fetzer, James H. 2001. *Computers and cognition: Why minds are not machines*. Vol. 25. Springer Science & Business Media.

Floridi, Luciano. 2019. "Translating principles into practices of digital ethics: Five risks of being unethical." *Philosophy & Technology* 32 (2): 185-193.

Gates, Edward. 1989. "Webster's new world dictionary." *English Today* 5 (2): 52-54.

Gunkel, David J. 2012. *The machine question: Critical perspectives on AI, robots, and ethics*. mit Press.

Hagendorff, Thilo. 2020. "The ethics of AI ethics: An evaluation of guidelines." *Minds and machines* 30 (1): 99-120.

Hosseini, Mohammad, David B Resnik, and Kristi Holmes. 2023. "The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts." *Research Ethics* 19 (4): 449-465.

Johnson, Robert, and Adam Cureton. 2004. "Kant's moral philosophy."

Marr, Bernard. 2024. *Generative AI in practice: 100+ amazing ways generative artificial intelligence is changing business and society*. John Wiley & Sons.

McKitterick, Alistair J. 2022. "The role of teleology in practical theology." In *Evangelicals Engaging in Practical Theology*, 56-69. Routledge.

Mittelstadt, Brent. 2019. "Principles alone cannot guarantee ethical AI." *Nature machine intelligence* 1 (11): 501-507.

Moor, James H. 2012. *The Turing test: the elusive standard of artificial intelligence*. Vol. 30. Springer Science & Business Media.

Morton, Timothy. 2016. *Dark ecology: For a logic of future coexistence*. Columbia University Press.

Müller, Vincent C. 2025. "Philosophy of AI: A structured overview." *A Companion to Applied Philosophy of AI*: 14-30.

Nowak, Andrzej, Paul Lukowicz, and Pawel Horodecki. 2018. "Assessing artificial intelligence for humanity: Will ai be the our biggest ever advance? or the biggest threat [opinion]." *IEEE Technology and Society Magazine* 37 (4): 26-34.

O'Regan, Gerard. 2008. "Artificial Intelligence and Expert Systems." In *A Brief History of Computing*, 149-177. Springer.

Palm, Elin, and Sven Ove Hansson. 2006. "The case for ethical technology assessment (eTA)." *Technological forecasting and social change* 73 (5): 543-558.

Ruttkamp-Bloem, Emma. 2020. "The quest for actionable AI ethics." Southern African Conference for Artificial Intelligence Research.

Segun, Samuel T. 2021. "From machine ethics to computational ethics." *Ai & Society* 36 (1): 263-276.

Tamuno Miegbam, Alex, and Kia Bariledum. 2022. "Artificial intelligence and diplomacy in the 21st Century: The African perspective." *Central Asian Journal of Theoretical and Applied Sciences* 3 (10): 49-65.

Turing, Alan M. 2009. *Computing machinery and intelligence*. Springer.

Veenhoven, Ruut. 2014. "Greatest Happiness for the Greatest Number." In *Encyclopedia of Quality of Life and Well-Being Research*, 2612-2613. Springer.

Wesseh, Charles W. 2025. "Utilitarianism Vs. Deontology: A Comparative Study of Ethical Theories in Moral Philosophy." *Available at SSRN 5222906*.