European Journal of Sustainable Development Research
2027, 11(1), em0443
e-ISSN: 2542-4742
https://www.ejosdr.com/

MODESTUM

Perspective Article

OPEN ACCESS

Artificial intelligence and SDGs, a friend or foe: A perspective of Kantianism

Elizabeth Lee [1*], Tan Seng Teck [1], Mohd Rushidi [1], Chee How Liau [2], Pradeep Sonar [3]

[1] INTI International University Malaysia, MALAYSIA
[2] Lecturer in Marketing, Danford Higher Education, Melbourne, AUSTRALIA
[3] Parul University Vadodara, Gujarat, INDIA
***Corresponding Author:** i25038231@student.newinti.edu.my

Citation: Lee, E., Teck, T. S., Rushidi, M., Liau, C. H., & Sonar, P. (2027). Artificial intelligence and SDGs, a friend or foe: A perspective of Kantianism. *European Journal of Sustainable Development Research, 11*(1), em0443. https://doi.org/10.29333/ejosdr/19643

ARTICLE INFO

Received: 17 Apr. 2026
Accepted: 16 Aug. 2026

ABSTRACT

Humans have keen interests with artificial objects. We consume food and drinks containing artificial coloring while FIFA approves the use of artificial turf for football matches. However, the rise of artificial intelligence (AI) has gripped the world with its radical changes, triggering not only physical, but also philosophical implications. In this paper, we explore the impact of AI on the Sustainable Development Goals (SDGs) and pose the question of whether the human centrality of the SDGs could diminish with the introduction of AI. Drawing on Kant's concept of universalism and categorical imperatives, we argue that AI cannot safeguard intergenerational justice or meet Kant's definition of a rational being. This paper presents a brief yet significant perspective, emphasizing the need for caution when integrating AI into the implementation of the SDGs, particularly concerning policy design.

Keywords: artificial intelligence, SDGs, Kantianism, economic growth

INTRODUCTION

There is a consensus among the global community that the world is facing serious environmental challenges, including pollution, climate change, and waste management. These issues are becoming increasingly severe due to rampant urbanization and industrialization (Talaie et al., 2024). Addressing these challenges requires innovative technological solutions, and Artificial Intelligence (AI) has emerged as a potential answer to these pressing environmental concerns. Expectations are growing that AI can drive radical change and accelerate the achievement of the Sustainable Development Goals (SDGs) by "harnessing the benefits of artificial intelligence and addressing the risks… to deliver on the promise of technological advances and harness them for the common good" (Guteres, 2019). AI can analyze vast amounts of data, make fact-based decisions, and provide predictive tools for various outcomes. However, the use of AI is not without ethical risks and may jeopardize the core philosophical value underlying the SDGs.

Leveraging Kantian ethics, this paper explores the compatibility of AI and the increasing automation of tasks with Kantianism, which emphasizes the importance of humanity. The discussion surrounding AI and Kantianism has been broad, particularly regarding the potential for "giving machines ethical principles or procedures to resolve the ethical dilemmas they might encounter, enabling them to function ethically through their own decision-making" (see Anderson & Anderson, 2011, p. i). However, Mana et al. (2021) argue that AI is not feasible from a Kantian perspective. They contend that Kant's duties must be executed through free will and happiness—elements that AI mechanization does not support. Consequently, they argue that AI should only serve as a peripheral utility. In contrast, Shetty (2025) posits that AI relies on human intervention, which, through evolution and adaptation, has developed the capacity to interpret the world—something AI cannot achieve independently. Shetty asserts that "AI, by contrast, as an artifact of human creation, lacks the independent capacity to redefine its boundaries in an open-ended manner" (Shetty, 2025, p. 4), which Baiasu (2022) describes as one the greatest problems in cognitive science.

These are valuable works, but they do not address the invocation and the implications of AI on SDGs. Our paper intends to fill this gap and explore the philosophical compatibility of AI in SDGs. This is a question of significance because SDGs emphasize the importance of the humanities and the critical examination of the human condition in understanding the moral significance and urgency of pursuing these goals; that is, "the instrumental value of the *humanities* in pursuit of the development goals with the assumption that the goals are themselves important and urgent. Yet our

Copyright © 2027 by Author/s and Licensed by Modestum DOO, Serbia. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

concern is *not only about successful pursuit of given goals but also about their value and importance* (Macleod, 2026) [Emphasis added]. Since SDGs place emphasis on the value of humanity, it then bags the question if the application of AI in SDGs implementation would could ultimately diminish humanity centralized in SDGs. Our work examines the controversial intersection—if there is—between the infiltration of AI into the SDGs and the notion of humanity, which lies at the core of these foundational principles.

We emphasize two essential premises of Kantianism. First, it provides a human-centered ethical framework that places human life and ability at the core of a norm-creating moral philosophy, shaping our understanding of moral behavior. Second, the goal of Kantian ethics is to offer a practical philosophy that is both useful and relevant to the pursuit of universal and categorically imperative moral behavior. This work is structured as follows: we first explore, briefly, the rise of AI and its implementation in SDGs. We then establish that Kantianism underpins the moral philosophy of the SDGs. We then examine the main tenets of Kantianism and analyze the extent to which AI fulfills or contradicts the Kantian requirements regarding the implementation of the SDGs. Towards the end, we propose how AI can be implemented guardedly so that it upholds humanity while improving the efficiency of SDG operations.

THE RISE OF AI AND SDGs

The late 1930s marked the beginning of the exploration of the concept of AI (Turing, 1950). Over the previous six decades, AI has made substantial progress since its inception in 1956. Significant academic and industrial interest has been shown in this matter. In the early phases of AI, McCulloch and Pitts' "brain-like machine" (1943) and Turing's earlier work "Computing Machinery and Artificial Intelligence" (Turing, 1950) were both significant proponents, which facilitated the gestation of AI during the 1950s. AI's advancement was further solidified during the 1960s as it captured the attention of academics, industry, and the medical faculty. During this period, the inaugural International Joint Conference on AI (JCAI) was convened in 1969. Large conglomerates, including IBM, were significantly interested in these events, which incited collaboration between industry and academia. This attention was particularly important during the initial stages of AI development.

Experts from other fields recognized the limitations of AI in the early 1970s, asserting that "artificial intelligence is not a scam, but it is also a mediocre self-interference." Consequently, AI development encounters obstacles. In the 1990s, the internet introduced a multitude of data that could be used to train models, resulting in a substantial change in the landscape. Nevertheless, AI technology was revitalized at this time. An expert system developed by IBM, Deep Blue, achieved a significant milestone in 1997 by defeating Garry Kasparov, the incumbent chess champion. This achievement serves as evidence of the extensive commercialization of AI. Using a meticulous evaluation and accounting of each conceivable maneuver, Deep Blue implemented a methodical brute-force strategy. Even though Deep Blue's victory over a human opponent was a significant milestone, it is essential to acknowledge that its accomplishments were limited to a specific domain: the chess rules. Modeling the profound intricacies of the broader world was substantially constrained by the system's capabilities.

From 2010 to the present, AI has made substantial progress, with the development of Generative Pre-Trained Transformers (GPTs) being a significant turning point in the field (Delipetrev et al., 2020). ChatGPT was developed by OpenAI, a company that has since developed a variety of GPTs. The first GPT was introduced in 2018. The most noteworthy of these innovations is ChatGPT. At first, it was determined that the public would not have access to GPT-2 and GPT-3. In June 2020, GPT-3 was ultimately introduced to the public, setting a new standard with an impressive 175 billion parameters. Consequently, it established itself as one of the most potent language models of its era. The GPT-3 system can generate text that is contextually pertinent, coherent, and human-like in response to input prompts. All its capabilities are included in this capability. In the realm of natural language processing, it can perform a diverse array of tasks, including the generation of poetry, the engagement in conversations, the composition of essays, and the answering of queries. Through OpenAI's Application Programming Interface (API), developers were granted access to the model, which allows them to integrate its language generation capabilities into their custom applications and services. ChatGPT, a chat-based interface that was developed on the GPT-3.5 language paradigm, was introduced by OpenAI in 2022.

Of course, AI has made significant contributions to the administration and monitoring of the 17 SDGs within the context of the SDGs. In the context of SDG 13, which prioritizes climate action, it is crucial to acknowledge that technological advancements in artificial intelligence are enhancing our understanding of climate change and the modeling of its potential repercussions. Additionally, artificial intelligence has the potential to accelerate the advancement of low-carbon energy systems, which are indispensable in the battle against climate change. The utilization of renewable resources by these systems improves energy efficiency. In addition, AI has the capacity to enhance the health of ecosystems by automating the detection of oil accidents, thereby preventing and substantially reducing marine pollution (Keramitsoglou et al. 2006). Artificial intelligence has also been extensively employed to mitigate deforestation and restore degraded land. The classification of desolate lands that are documented in satellite imagery is facilitated more rapidly and effectively (Mohamadi & Kavian, 2015). The identification of large-scale patterns of desertification and the improved allocation of resources is facilitated by this capability. Further, it provides valuable information for environmental management, planning, and decision-making, which can be employed to prevent further desertification or rectify existing trends by identifying the underlying causes (Chaudhary, 2023) and to improve biodiversity conservation (Kwok, 2019).

However, the contributions of AI in SDGs are not without controversy. For instance, AI has significantly contributed to precision farming, a management strategy that enhances the efficiency, productivity, quality, profitability, and

sustainability of agricultural production by collecting, processing, and analyzing temporal, spatial, and individual data. This technology is valuable for predicting erratic weather, which is essential for farmers to proactively prepare for potential contingencies, such as droughts or food shortages. However, this tool can serve as a double-edged sword; poorer farmers may find it ineffective, as it does little to alleviate their challenges. Additionally, it has enabled governments and banks to withhold loans from these individuals if bad weather and poor harvests are forecasted (Stahl et al., 2023).

The use of AI has also raised concerns about civil liberties infringements. For example, the 'Helicopter Research' initiative allowed researchers to track the movements of mobile phone users in Sierra Leone, Liberia, and other parts of Africa affected by the Ebola outbreak. This tracking was conducted with the belief that it could help monitor the outbreak; however, no ethical guidelines were established to limit its extent. As a result, the tracking was executed haphazardly, and findings revealed that those monitored were aware of the travel ban, with many choosing not to travel to the outbreak areas. Unfortunately, valuable time was spent debating whether to use the tool, a discussion many felt would have been better directed toward more substantive measures to curtail the outbreak (Maxmen, 2019).

KANTIANISM AND SDGs

We have briefly discussed the rise of AI and the increasing integration of its mechanisms into the SDGs. We must now justify why we believe that Kantianism provides an appropriate foundation for the moral philosophies underpinning the SDGs. This discussion is of essence because humans are now in the era of the Anthropocene and important ecological crossroads. Identifying the necessary moral philosophy underpinning this new Enlightenment project is essential as pandemics and ecological catastrophes loom large on the horizon. Taking a wrong philosophical approach to implementing AI in this fragile context would threaten both moral and political paralysis (Vereb, 2023). However, there is no unison of thoughts on the underpinning moral philosophy in AI adoption for SDGs. But there is also no room to engage in an extensive debate in this paper. For this work, we claim that Kantianism provides a satisfactory account of guiding principles on AI adoption. The reason, as we will show in the later parts of this work, is that AI is *antithetical* to humanity, as it leads to the existential question of what it means to be human in the context of a rapidly expanding technological age (Nourbakhsh & Keating, 2020). That is "How the development of AI affects the cognitive capacities of humanity will depend on which route humanity adopts for the development of AI. If AI develops in a way in which its functioning in higher cognitive tasks is based on knowledge inaccessible to humans, and the main goal of AI is to *replace humans in tasks requiring complex cognition*, then the *consequences may be disastrous*" (Nowak et al., 2018, p. 34).

This paper insists that Kantian offers a rigid defense to humanity, and the invocation of AI using Kantian equation would instill a more balanced adoption of AI and achieving an envisaged balance of efficiency and preserving human dignity. However, placing Kantianism at the center of environmental reform is not without its challenges. One significant obstacle is Kant's seemingly *non-collective* approach to environmental ethics, which prioritizes human welfare as its primary concern. Kant's writings assert:

> The fact that the human being can have the 'I' in his representations raises him infinitely above all other living beings on earth. Because of this he is a person, and by virtue of the unity of consciousness through all changes that happen to him, one and the same person – i.e., through rank and dignity an entirely different being from things, such as irrational animals, with which one can do as one likes. (Louden & Kuehn, 2012, p. 15)

Kant's prioritization of human interests over all else is problematic and has created enduring issues, even though his later, more environmentally conscious views appear to be insufficient to alter this detrimental perception. It has been acknowledged by scholars such as Rolston and Garvey that Kant's non-anthropogenic posture renders him an unsuitable candidate for advocating for critical environmental ethics reforms. According to Garvey (2008), environmental ethics is primarily concerned with broadening our understanding of value, or at least the number of things we value, in contrast to Kant's restrictive perspective that places humans above all else (Garvey, 2008, p. 51). Broome (2012) argues that the environmental crisis is best addressed through a utilitarian framework, which is countered by Kant's restrictive approach to environmental matters. For Broome, the utilitarian cost-benefit equation enables an individual and society to make a comprehensive decision regarding which actions must be sacrificed and which actions must be committed to advance the public welfare. Broome (2012) posits that the economics of utilitarianism provides a value-free approach to protect the most important interests of society.

We contend that many criticisms of Kant are largely unfounded and we dedicate this section to the defense of Kant. First, Kant's focus on humans as rational beings does not render non-rational beings irrelevant. It does not suggest that humans are free to exploit animals, clear forests without restraint, or consume all land animals indiscriminately. Kant argued that in the pursuit of universal happiness, humans must treat others as any other rational beings would. His writings assert:

> ...if a man shoots his dog because the animal is no longer capable of service, he does not fail in his duty to the dog, for the dog cannot judge, but his act is inhuman and *damages in himself that humanity which it is his duty to show towards mankind*. If he is not to stifle his human feelings, he must practice kindness towards animals, for he who is cruel to animals becomes hard also in his dealings with men (Lecture on Ethics, 212 (27: 459) [Emphasis added].

In another lecture, Kant fortifies his position:

> A propensity to wanton *destruction of what is beautiful in inanimate nature* (spiritus destructionis) is *opposed to a human being's duty to himself*; for it weakens or uproots that feeling in him which, though not itself moral, is still a disposition [Stimmung] of sensibility that greatly promotes morality or at least prepares the way for it: the disposition, namely, to love something (e. g., beautiful crystal formations, the indescribable beauty of plants) even apart from any intention to use it. (…) (Svoboda, 2012, p. 100) [Emphasis added].

The two excerpts exemplify Kant's emphasis on developing humanity rather than individuals. Kant's transcendental ideas instill a sense of transgenerational justice that encompasses not only the common good of the present generation but also that of future generations. This transcendental perspective has a much broader impact on the community compared to the utilitarian approach to environmental ethics. For Kant, this pursuit represents a continual "cultural project of enlightenment," which he describes as "the exit of human beings from their self-incurred immaturity…Dare to know!…Have the courage to use your own reason!" Enlightenment cannot be bestowed upon us by others; it requires "using our own reason" (Fleischacker, 2015). Thus, Kant's formulation emphasizes on human dignity, that is we are obliged to "act that you use humanity, whether in your own person or in any other person, always at the same time as an end, never merely as a means" (Kant, 1998, p. 66). This principle underlies Kant's belief that individuals and, indeed, the world should not be exploited for personal or societal gain. In his view of environmental ethics, each generation has a responsibility to safeguard the environment for those that follow. The emergence of governmental paternalism exemplifies the protection of intergenerational justice, establishing a universal standard to shield future generations from harm and exploitation. The recent implementation of the Tobacco Plain Packaging Act 2011 (Cth) in Australia serves as a strong illustration of Kant's concept of intergenerational justice. The following excerpt highlights the transgenerational justice inherent in an Enlightenment project:

> …tobacco, says the World Health Organization (WHO) "is the only legal consumer product that kills when used exactly as intended by the manufacturer." (Rimmer 2013). With overwhelming research showing that tobacco causes 85% of lung cancers (Price et al. 2023), the Australian government upscaled their paternalism and boldly imposed the first-ever plain packaging policy in the world. The Tobacco Plain Packaging Act 2011 (Cth) of Australia governs the retail packaging and aesthetic of tobacco products, mandating unadorned, olive-hued packaging featuring public health warnings and graphic depictions of smoking-related illnesses. The Australian government, progressing towards universal justice and…chose to enact a universal law, despite the potential withdrawal of investments by tobacco companies (Teck, 2025).

Kant's concept of imperative duty and universal law creates an internal metronome of "metaphysics of responsibility" (Schönfeld & Lee, 2010), which ought to be fulfilled by the rational being, taking the primal responsibility to ensure the achievement of common good for all. His moral duty, while valuing rational beings more than non-rational ones, ultimately establishes a cosmopolitan perspective focused on the ends and happiness for all (Vereb, 2022). By framing the duty of an agent as one that commits to actions conforming to universal law, and through his insistence on categorically imperative decisions, Kant emphasizes the importance of developing a collective ideal that encompasses intergenerational justice. This collective notion is further reflected in his focus on achieving a collective end rather than treating it as merely a means to an end. The criticism that Kant neglects non-rational beings can be simplistically defended by asserting that only rational beings possess the transcendental freedom necessary to make reasoned judgments, a capacity that is evidently lacking in lower-ranked beings.

Kant presents another support towards environmental ethics as he encourages the forging of international collaborations through laws and regulations, a notion that resonates with SDGs urgent call for action by all countries - developed and developing - in a *global partnership*. "Kant's views on hope and political cooperation underscore … humanity's overarching moral vocation, which, in light of climate change, can be directed toward ecological stewardship" (Vereb, 2022). Kant invites stewardship, a collective endeavor without which "humanity will likely exacerbate climate impacts affecting others across the globe" (Vereb, 2022), which resonates with UN's call for all nation-states to achieve the 17 designated SDGs. Kant's cosmopolitan stewardship is best summarized as follows:

> Kant is suggesting that any state constitution and civil order, no matter how internally coherent and stable, cannot be fully secure unless its external relationships with other states are also mutually secure and that this can only be done through meaningful political cooperation and the acceptance of genuine responsibilities between states. As a result, it is here, in the need to promote greater responsibilities between states (so as to secure domestic security) that we can see that Kant provides a basis for what might be called transitional cosmopolitanism (Brown et al., 2019 p.4).

AI, SDGs AND THE ESSENTIALS OF KANTIANISM

We have argued that Kantianism provides a foundational moral theory for the SDGs. Now, we turn to the question of whether the use of AI in achieving these goals aligns with Kantian moral imperatives. It is important to note that the UN has recognized the potential risks associated with employing AI in the context of the SDGs. They stated:

> Although there is tremendous potential for AI, *it is not without risks*. While AI can drive progress, it also brings *serious challenges, especially when used in the absence of clear ethical guidelines*. When working with AI to solve some of our world's toughest challenges, it's imperative to consider: Bias and discrimination: AI models trained

> on non-representative data reinforce inequalities in access to services or employment. Surveillance and privacy: The use of AI in monitoring must be balanced with respect for human rights. Labour market disruption: Automation may disproportionately affect low-income and marginalized workers, particularly in the Global South. Environmental impact: The energy use of large AI models significantly contributes to carbon emissions if not managed sustainably... Artificial Intelligence is not a silver bullet—but it is a catalyst. With thoughtful implementation, AI can dramatically speed up our transition to a just, inclusive and sustainable world The challenge lies not only in what AI can do but in what we, as a global community, choose to do with it (United Nations Global Compact, 2025) [Emphasis added].

To answer the philosophical question if AI in SDGs collides with Kantianism, we will first need to examine the essentials of Kantian theory. The main impression of Kantianism is its utmost emphasis on a human-centered moral framework, and that his universal law implies a three-pronged requirement of "categorical imperative, autonomy of will, and the *dignity of humanity* as the end in itself" (Ulgen, 2017) [Emphasis added]. We will now peruse the underlying theorem of Kantianism and examine to what extent Kant's views are parallel to the use of AI in SDGs.

AI and Human Dignity

Kant emphasizes the necessity of establishing a set of categorically imperative rules that form the foundation of every moral judgment, articulated as "if the action is represented as good in itself and therefore as necessary" (Kant, 1997). This concept essentially describes a rule with intrinsic consequences that is objective and not influenced by intended results. As Kant states, "but with its form and with the principle from which it follows; and what is essentially good in the action consists in the mental disposition, let the consequences be what they may" (Kant, 1997). This rule requires agents to follow a doable, intelligible, and rational constraint that promotes the welfare of others, ensuring that individuals are not treated merely as means to an end (Ulgen, 2017). Ulgen was particularly skeptical if ever machine-mediated human interaction can coincide with ethical conduct and humanity, especially if it would be desirable to replace human dignity with these technologies. Ulgen argues that the process of replacing human dignity with machines is not fundamentally beneficial to humankind because machines are unable to generate a 'will' of their own, nor have the capacity to set rules and abide by them. This rule-making limitation severely hampers the human autonomy which Kant thought is imperative for rational beings. Kant argues that an action that cannot be universally adhered to or one that an individual would not willfully follow cannot be considered a categorical imperative, as it undermines the autonomy of other agents. In summary, establishing the scene for the categorically imperative and universal rule necessitates a rational framework of doable, intelligible, and attainable rules for agents that collectively yield a beneficial outcome for the common good. The pertinent question is whether the use of AI aligns with this rule. This alignment depends on whether a supplementary or arbitrary approach is adopted (see Ulgen, 2017).

We will begin by examining the arbitrary autonomous AI approach. Some scholars contend that "a rule-based ethical theory is a good candidate for the practical reasoning of machine ethics because it generates duties or rules for action, and rules are (for the most part) computationally" (see Powers, 2006 p. 46-47). Powers (2006) argues that a "rule-based ethical theory is a good candidate for the practical reasoning of machine ethics because it generates duties or rules for action, and rules are (for the most part) computationally tractable" (p. 47). However, this paper presents two objections to this approach. First, Kant stresses the idea of choice, transcendental freedom to rationalize what is good not only for oneself but good to those who are parallelly impacted by those actions. In Kant's words:

> ...it is this *transcendental idea of freedom* on which the practical concept of freedom is grounded, and the former constitutes the real moment of the difficulties in the latter, which have long surrounded the question of its possibility. Freedom in the practical sense is the independence of the power of choice from necessitation by impulses of sensibility. For a power of choice is sensible insofar as it is pathologically affected (through moving-causes of sensibility); it is called an animal power of choice (arbitrium brutum) if it can be pathologically necessitated. The human power of choice is indeed an arbitrium sensitivum, yet not brutum but liberum, because sensibility does not render its action necessary, but in the human being there is a faculty of determining oneself from oneself, independently by sensible impulses...it is easy to see if all causality in the world of sense were mere nature, then every occurrence would be *determined in time by another in accord with necessary laws*, and hence – since appearances, insofar as they determine the power of choice, would have to render every action necessary as their natural consequence – the *abolition of transcendental freedom would also simultaneously eliminate all practical freedom*. For the latter presupposes that although something has not happened, it nevertheless ought to have happened, and its cause in appearance was... (Anamarie, 2022 p 22-23) [Emphasis added].

These passages reflect Kant's firm rejection of a pervasive approach to AI. The concept of transcendental freedom suggests that the moral agent has an ongoing duty that transcends temporal limits, a point Powers concedes. "I've suggested three accounts, according to which we might conceive of a deontological ethical machine. Each account has its challenges—triviality, asymmetry, excessive specificity, lack of semidecidability, and lack of priority for maxims, to repeat those I've described here. Although these problems seem difficult to surmount, they are similar to problems in human attempts to engage in practical reasoning" (p. 51). This duty involves not only making the best unbiased decision relevant to the present time but also making a necessary decision that offers parallel benefits to other agents across different times and spaces. Kant's ideas about happiness and

his descriptions of the 'holy' or 'divine' will support this assertion. Kant's categorical imperative requires humans to continuously strive for categorical rather than hypothetical imperative, that is an imperative that produces happiness to conform to natural law, instead of happiness that conforms to extraneous conditions. In his words:

> Happiness is the satisfaction of all of our inclinations (extensive, with regard to their manifoldness, as well as intensive, with regard to a degree, and also potensive, with regard to duration). The practical law from the motive of happiness I call pragmatic (rule of prudence), but that which is such that it has no other motive than the worthiness to be happy I call moral (moral law). The first advises us what to do if we want to partake of happiness; the second commands how we should behave in order even to be worthy of happiness (Kant, 1997).

We now return to our point regarding whether fully autonomous AI would aid in the attainment of natural happiness or the divine will. AI technology, even at its most advanced stage, relies on effective programming and the incorporation of data sets, which means that its ability to contribute to natural happiness or divine will is contingent upon the ethical frameworks and values embedded within those programs. Scientists predict that the future of AI may bring about a system capable of self-programming, which Kurzweil (2005, p. 17) describes as "an intelligent explosion, which will be powerful enough to replicate human moral agency as well as the brain itself." The question is whether AI, in its most developed form, could fulfill Kant's call for the fulfillment of natural happiness. We argue that AI, even in this advanced state, cannot meet Kantian requirements.

First, we contend that an agent's decision-making in transcendental freedom requires more than cognitive function; it necessitates a rational choice that embeds humanity at the core of the decision-making process. AI, no matter how advanced it is, is programmed by people to try to make the programmer's idea of happiness come true. This perspective aligns with the assertion that "the determinant of 'will' is the 'happiness or pleasure' of human users" (Johnson 2006: 202). Second, we maintain that decisions based solely on statistical data cannot fulfill the need for transcendental decisions that extend beyond temporal limits, thus rendering the achievement of intergenerational ecological justice seemingly unfeasible (Farnsworth 2017). This limitation highlights the necessity for ethical considerations and long-term thinking in policymaking to ensure sustainable outcomes for future generations.

The remaining query then is whether semi-autonomous AI would satisfy Kant's philosophical requirements? This too in our opinion is not difficult to refute because Kant requires *absolute* adherence to natural law, not partial. We refer to this passage, "Kant claimed that if someone is *solely subjected* to transcendental causality, then s/he would not achieve the option to perform otherwise. Kant has denominated them as the 'holy will'. Their actions will necessarily follow the transcendental causality; hence there is *no 'choice'* to *execute otherwise*… the notion of 'imperative' in this regard" (Manna & Nath, 2021).

AI and Autonomy

We now turn to the element of autonomy, which Kant defines as "the property the will has of being a law to itself (independently of every property belonging to the object of volition)" (Kant, 1997). The autonomy of will is a spillover condition stemming from the transcendental freedom of an agent, allowing for decision-making free from external influence, guided by a duty to fulfill the rules of natural happiness. Proponents of AI often argue that it embodies an unbiased and objective commitment to following rules and laws, as it operates without external influence and appears to generate decisions with plausible objectivity (Bendel, 2019).

We contend that advocates of this perspective are superficial and misguided. In his work "Groundwork for the Metaphysics of Morals," Kant distinguishes among immoral, amoral, and moral actions. Moral actions align with the duty to comply with natural laws, while immoral actions do the opposite. Amoral actions, which AI is deemed capable of performing, are those made in compliance with the duties imposed by the programmer rather than through a transcendental obligation. From this standpoint, we argue that AI's seemingly amoral decisions, which adhere to the fundamentals of programming, reflect at best an amoral disposition—an adherence to the duties of humans rather than duties to nature (Manna & Nath, 2021).

Our previous arguments suggest that AI is a poor candidate to fit Kantian terminology regarding a moral agent. AI lacks the ability to think from a universal perspective. The capacity for universal thought and the understanding of what is categorically imperative involve more than just cognition; they require transcendental freedom, reasoning, and enlightenment. It remains questionable whether AI, even in its most advanced form, can achieve this highest level of reasoning.

ACHIEVING A BALANCE

The question we now ask is how to achieve a balance between leveraging AI and maintaining the integrity of Kantian values within the SDGs. To facilitate our discussion, we will divide our suggestions into three segments. Drawing on the essentials of Kantianism, we propose that autonomy must be preserved at every level of AI integration into the SDGs. This means that at each level and for each purpose, the objectives of AI should align with the UN Charter and the principles of the SDGs. Implementing AI in parallel with the UN Charter can provide a safeguard against misguided objectives while still adhering to the overarching perspectives of the charter. AI should never be permitted to make policy decisions on behalf of humans; its role should remain limited to facilitation and operational aspects (Mohammad et al., 2025). We argue that while AI can assist in the implementation of the SDGs, the final decision-making autonomy must rest with humans. In this regard, we support the proposal that "specific socio-technical design patterns—including careful role specification, implementation of defeater mechanisms, and support for reflective practice—can help maintain domain-specific autonomy while leveraging AI capabilities" (Bermudez, 2025 p.1). One method is the proposed use of

integrating self-awareness as a protective measure against value violations. Self-knowledge is a term that is used in the broadest sense to refer to the awareness of one's emotions, thoughts, beliefs, desires, and values (Gertler, 2010). For instance, Carter (2022) proposed the inclusion of the design friction for data privacy decision-making to support decision-makers to make more reflective, value and policy centered decisions in any complex conditions. Chen & Schmidt (2024) advocated the use of positive friction method for a collective and more harmonious collective decision making and as well as a reflective tool. These are technical jargons, which are beyond the scope of this paper to address and are left in the hands of qualified specialists. The ultimatum then is, whatever the outcome of innovations and software development, the goal is to maintain and preserve human dignity in the entire decision-making process and chain of command.

Second, the design of the AI systems must be overseen by an independent committee that comprises of a conglomeration of different stakeholders, that are enabled to provide views in co-designing the system. This allows a more balanced and holistic approach to ensure that the overall happiness of humanity is achieved and adequately preserved. Leveraging on the works of Chong et al (2023), we argue that the social and proxy agency provides a good tension and asymmetrical cocreation into designing an AI system that "fits" all. This model offers a tensioned focused model of co-creating an AI system that considers the welfare of all. "To address this challenge, we frame interactions between customers and embodied conversational agents through the lens of stakeholder engagement and apply the concept of proxy agency from social cognitive theory. This framework allows us to identify two primary stakeholder roles for embodied conversational agents: partner and servant. We conceptualize how these roles inform optimal design for embodied conversational agents and shape a two-stage value-by-proxy process, comprising proxy efficacy and outcome expectancy. Additionally, we uncover tensions within this process due to over-reliance on AI, as well as significant outcomes that extend beyond the immediate interaction" (Chong et al., 2023 p. 1).

UNCLAIMED PROVINCE

This paper is not all-encompassing. The imposition of AI and SDGs is enormous and includes broad philosophical discussions. Of course, the discussions on Kantianism and human dignity become futile if one begins to measure the values hedonically. Applying a hedonistic calculus would yield a totally different domain, that is, arguments then shift from ethics, dignity, and humanity to arguments that are cognitively and economically driven "based on functional and instrumental goals, reflect functional benefits, and are assessed on the basis of the degree to which a product is a means to an end" (Longoni & Cian, 2022). Utilitarianism is a powerful province, and measuring the advancements of AI in this form would yield a significantly different outcome. As we write, AI is replacing thousands of workers with more cost-effective and efficient alternatives. Meta CEO Mark Zuckerberg told employees in a memo Wednesday that the company's decision to lay off 8,000 employees is necessary because "success isn't a given" in the fierce and competitive space of artificial intelligence. In the memo, which CNBC viewed, Zuckerberg stated, "AI is the most consequential technology of our lifetimes." "The companies that lead the way will define the next generation" (Vanian, 2026). And on 20 May 2026, Puan Halimah Yacob, the former president of Singapore, slams Standard Chartered Bank CEO for describing the layoffs of its 15% of its workforce as not for "not cost-cutting…replacing in some cases lower-value human capital with the financial capital and the investment capital we're putting in" (Ler, 2026). Puan Halimah said, "Workers are human beings with families, not just a form of capital. They too have contributed to the bank and now, because of AI, have become redundant. It's demeaning to describe them as "lower-value human capital" (Ler, 2026).

Whether or not the replacing of workers is for cost cutting or whether they are inevitable trends of forwarding technology, one would question if AI has reduced human dignity to merely dispensable means. Whether one accepts the province of Kantianism on AI or utilitarianism is again a contest with a clear outcome. Utilitarianism has been the dominant theory in almost every aspect of human society, and we think AI is no exception (MacIntyre, 2013). Organizations, whether for profit or otherwise, would likely succumb to the ideals of utilitarianism, which provides a simple pain and pleasure calculus, and the utopian sensation that all is well if the interests of the majority are maximized. Hence, taking this route, as large corporations have already indicated, would justify the degradation of human dignity in the name of a moral simulacrum, an ideal that treats humans as a means to an end rather than the end.

CONCLUSION

We conclude our discussions with a cautionary statement from the late Professor Stephen Hawking: "The development of full artificial intelligence could spell the end of the human race … It would take off on its own and redesign itself at an ever-increasing rate" (Cellan-Jones, 2014). The Wall Street Journal reported in June 2025 that AI is learning to escape human control as it rewrote its own code to avoid shutting down. We cannot overlook these warnings. If we are correct in placing Kantianism as the central philosophical foundation of the SDGs, the imminent threat of AI overtaking humanity is disturbingly real. Kantian philosophy provides a rich anthropological context for the SDGs, and we have argued that Kant's concepts of happiness and universalism instill in the SDGs a cosmopolitan vision of collective interests. This understanding is crucial, yet the rise of AI in every aspect of life—ranging from automated gates to drones employed by Iran in asymmetric warfare with the USA—has generated a profound fear that the very foundation of human values, which the SDGs aim to protect, may be governed and managed by non-human entities. We must resist this outcome; otherwise, the SDGs may be left with a vacuum of philosophical ideal.

Author contributions: EL: writing – original draft; **TST:** conceptualization; MR: supervision; **PS & LCH:** writing – review & editing. All authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that this study is a conceptual study and does not require a formal ethical approval.

AI statement: The authors stated that generative AI tools was used to enhance linguistic fluency.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from corresponding author.

REFERENCES

Anderson, M., & Anderson, S. L. (Eds.). (2011). *Machine ethics*. Cambridge University Press.

Baiasu, S. (2022). The challenge of (self-)consciousness: Kant, artificial intelligence and sense-making. In H. Kim & D. Schönecker (Eds.), *Kant and artificial intelligence* (pp. 105-128). De Gruyter. https://doi.org/10.1515/9783110706611-003

Bendel, O. (2019). Are robot tax, basic income or basic property solutions to the social problems of automation? In *Proceedings of the AAAI 2019 Spring Symposium Series*. CEUR-WS. https://doi.org/10.26041/fhnw-6661

Broome, J. (2012). *Climate matters: Ethics in a warming world*. W. W. Norton & Company.

Brown, G. W., & Jarvis, S. (2019). Motivating cosmopolitanism and the responsibility for the health of others. In R. Beardsworth, G. W. Brown, & R. Shapcott (Eds.), *The state and cosmopolitan responsibilities* (pp. 203-223). Oxford University Press. https://doi.org/10.1093/oso/9780198800613.003.0011

Carter, S. E. (2022). A value-centered exploration of data privacy and personalized privacy assistants. *Digital Society, 1*(3), Article 27. https://doi.org/10.1007/s44206-022-00028-w

Cellan-Jones, R. (2014, December 2). Stephen Hawking warns artificial intelligence could end mankind. *BBC News*. https://www.bbc.com/news/technology-30290540

Chaudhary, G. (2023). Environmental sustainability: Can artificial intelligence be an enabler for SDGs? *Nature Environment and Pollution Technology, 22*(3), 1411-1420. https://doi.org/10.46488/nept.2023.v22i03.027

Chong, T., Yu, T., Keeling, D. I., de Ruyter, K., & Hilken, T. (2026). Stakeholder engagement with AI service interactions. *Journal of Product Innovation Management, 43*(1), 31-56. https://doi.org/10.1111/jpim.12786

Delipetrev, B., Tsinaraki, C., & Kostić, U. (2020). *Historical evolution of artificial intelligence* (EUR 30221 EN). Publications Office of the European Union. https://doi.org/10.2760/801580

Fleischacker, S. (2015). Kant's Enlightenment. *Con-Textos Kantianos: International Journal of Philosophy, 2*, 177-196.

Garvey, J. (2008). *The ethics of climate change: Right and wrong in a warming world*. Continuum.

Gertler, B. (2010). *Self-knowledge*. Routledge.

Kant, I. (1997). *Lectures on ethics* (P. Heath & J. B. Schneewind, Eds.). Cambridge University Press.

Keramitsoglou, I., Kiranoudis, C. T., Ceriola, G., Weng, Q., & Rajasekar, U. (2011). Identification and analysis of urban surface temperature patterns in Greater Athens, Greece, using MODIS imagery. *Remote Sensing of Environment, 115*(12), 3080-3090. https://doi.org/10.1016/j.rse.2011.06.014

Kwok, R. (2019). AI empowers conservation biology. *Nature, 567*(7746), 133-134. https://doi.org/10.1038/d41586-019-00746-1

Ler, S. (2026, May 20). 'Disturbing': Halimah Yacob says StanChart CEO calling retrenched staff 'lower-value human capital' demeaning. *AsiaOne*. https://www.asiaone.com/singapore/halimah-yacob-standard-chartered-retrenchments-artificial-intelligence-ai-disturbing-demeaning

Longoni, C., & Cian, L. (2022). Artificial intelligence in utilitarian vs. hedonic contexts: The "word-of-machine" effect. *Journal of Marketing, 86*(1), 91-108.

MacIntyre, A. (2013). *After virtue*. Bloomsbury Academic.

Macleod, C. M. (2026). 'It takes a village' or why the humanities matter to the SDGs. *Association of Commonwealth Universities*. https://www.acu.ac.uk/the-acu-review/it-takes-a-village/

Manna, R., & Nath, R. (2021). Kantian moral agency and the ethics of artificial intelligence. *Problemos, 100*, 139-151. https://doi.org/10.15388/Problemos.100.11

Maxmen, A. (2019, May 29). Can tracking people through phone-call data improve lives? *Nature*.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*(4), 115-133.

Mohamadi, M. A., & Kavian, A. (2015). Effects of rainfall patterns on runoff and soil erosion in field plots. *International Soil and Water Conservation Research, 3*(4), 273-281. https://doi.org/10.1016/j.iswcr.2015.10.001

Mohammad, S. I., Yogeesh, N., Raja, N., William, P., Ramesha, M. S., & Vasudevan, A. (2025). Integrating AI and fuzzy systems to enhance education equity. *Applied Mathematics & Information Sciences, 19*(2), 403-422. https://doi.org/10.18576/amis/190215

Nourbakhsh, I. R., & Keating, J. (2020). *AI and humanity*. MIT Press.

Nowak, A., Lukowicz, P., & Horodecki, P. (2018). Assessing artificial intelligence for humanity: Will AI be our biggest ever advance? Or the biggest threat? *IEEE Technology and Society Magazine, 37*(4), 26-34.

Powers, T. M. (2006). Prospects for a Kantian machine. *IEEE Intelligent Systems, 21*(4), 46-51. https://doi.org/10.1109/MIS.2006.77

Schönfeld, M., & Lee, J. (2010). The metaphysics of sustainability: Kant's categorical imperative. In *Sustainability and health* (pp. 1-18).

Shetty, R. (2025). Conscious limits: A Kantian perspective on the limits of human understanding and artificial intelligence. *Discover Artificial Intelligence, 5*, Article 1. https://doi.org/10.1007/s44163-025-00221-z

Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). AI for good and the SDGs. In *Ethics of artificial intelligence: Case studies and options for addressing ethical challenges* (pp. 95-106). Springer. https://doi.org/10.1007/978-3-031-17040-9_8

Svoboda, T. (2012). *Duties regarding nature: A Kantian approach to environmental ethics*. Pennsylvania State University Press.

Talaie, A., Kamyab, H., Chelliapan, S., & Khalili, E. (2024). Sources and impacts of dust pollution in Iran: A comprehensive overview. *Journal of Environmental Treatment Techniques, 12*(2), 1-15.

Teck, T. S. (2025). The trinity of corporate moral development: Mapping and extending "After Virtue." *Philosophy of Management, 24*(4), 433-453. https://doi.org/10.1007/s40926-025-00342-6

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

Ulgen, O. (2017). Kantian ethics in the age of artificial intelligence and robotics. *Questions of International Law, 43*, 59-83.

United Nations Global Compact. (2025, April 30). *Artificial intelligence and the Sustainable Development Goals: Operationalizing technology for a sustainable future*. https://unglobalcompact.org/compactjournal/artificial-intelligence-and-sustainable-development-goals-operationalizing

Vanian, J. (2026, May 20). Zuckerberg's Meta layoffs memo: 'Success isn't a given' in the AI era. *CNBC*. https://www.cnbc.com/2026/05/20/meta-layoffs-zuckerberg-says-success-isnt-a-given-in-memo.html

Vereb, Z. (2022). Sustaining the individual in the collective: A Kantian perspective for a sustainable world. *Kantian Review, 27*(3), 405-420.

Vereb, Z. (2023). A Kantian philosophy of hope for the 21st century? *Studi Kantiani, 36*, 111-126.