Received: 19 July 2022 | Revised: 20 March 2023 | Accepted: 19 April 2023

DOI: 10.1111/beer.12547

Business Ethics, the Environment & Responsibility

WILEY

PERSPECTIVE

All about the human: A Buddhist take on AI ethics

Chien-Te Lin

Institute of Religion and Humanities, Buddhist Tzu-Chi University, Hualien, Taiwan

Correspondence
Chien-Te Lin, Institute of Religion and Humanities, Buddhist Tzu-Chi University, Hualien, Taiwan.
Email: bhadanta@gmail.com, mahavirya@gms.tcu.edu.tw

Abstract

As AI technology becomes more influential, ethical considerations surrounding its application are becoming increasingly relevant. In this paper, I reflect on some moral questions from a Buddhist perspective and consider the moral status of AI to evaluate its function and purpose in our lives. Since a robot lacks the capacity to experience suffering and has no conscience, AI ethics are possible only as ethics *about* robots and not as ethics *for* robots. Despite having no concrete moral status, robots cannot be deemed entirely morally insignificant as they exist as moral objects toward which moral agents have a duty. Our unique endowment as human beings is the intelligence for moral deliberation as we develop AI technologies and determine the future direction of humanity. From a Buddhist point of view, there is potential to advance our spiritual growth through the realization of the interdependence between humans and AI. As we recognize the uniqueness of our humanity and take a middle-way approach, the rise of robots need not threaten our existence but could instead catapult humanity into a new dawn.

KEYWORDS
AI ethics, Buddhism, common humanity, consciousness and conscience, the theories of dependent origination and middle way

1 | INTRODUCTION

Through recent advances in science and technology, we have all been introduced to "Sophia"—an Artificial Intelligence program that near-flawlessly interacts with humans. Indeed, AI has emerged as a dominant force in various spheres of life.[1] Far-reaching AI technology is already being implemented in economics in applications that aim to predict market trends and currency fluctuations. The logistics and automotive industries are experimenting with delivery drones and self-driving cars, and the influence of AI is extending into fields such as manufacturing, medicine, and law. But the same knife cuts both bread and fingers. Military inventions based on this technology, such as autonomous weapons or "slaughter-bots," have recently become hot topics of discussion (Kessel, 2019). While there is no doubt that such applications have the potential to inflict serious harm on human subjects through warfare, humans also frequently bring harm to AI robots. The threat of AI to job security has already caused global outrage, and attacks on robots seem to be emerging in an endless stream (Bromwich, 2019).

With AI technology set to dominate most aspects of modern society in the near future, it is becoming increasingly important to evaluate the moral implications of such advances in technology. As such, a number of issues are worth discussing, including: Can and should AI robots be programmed with moral codes? How do we view the loss of human life caused by robots, and the destruction of robots stemming from human activity? Is there a valid case for classifying AI applications as a being that requires the same sort of protection as humans, animals, or plants? Do AI beings have rights and/or duties? Or, more specifically, do killer robots bear responsibility for killing? How do we determine the moral status of an AI being?

As reflections on how robots should be treated and how they should treat others seem to rise in prevalence, the topic of roboethics is fast gaining attention in the Western world (Krishnan, 2009; Lin & Abney, 2014; Wallach & Allen, 2010; Tzafestas, 2016). However, such matters have largely remained under the radar as far as Eastern

 © 2023 John Wiley & Sons Ltd.

philosophy is concerned. Although Buddhist views on human rights and animal rights have been discussed by some (Keown et al., 1998; Waldau, 2018), AI ethics remains to be probed from a Buddhist perspective.[2] In this paper, I first aim to show that AI lacks moral accountability by virtue of not possessing consciousness and not having the capacity to experience suffering. Because of this, it can be argued that robots lack conscience and that, consequently, moral terms could not be applied to them. Second, I argue that although AI lacks status as a moral agent, this does not mean that artificially intelligent forms of being can be treated arbitrarily. From a Buddhist perspective, our unique endowments as human beings require that we not only develop AI from the goodness of humanity but also use it as a means to examine human morality and relieve suffering in this world, potentially assisting the pursuit of Bodhisattva practice. Finally, I apply the Buddhist teaching of the middle way to the issue of AI ethics and its sustainable development.

The contribution and novelty of this article is that it applies ancient oriental wisdom to contemporary AI issues. Instead of relying on a particular religious ideology, in this paper I show how the humanistic features of Buddhism reconfirm the unique role of humankind in this digital age, applying such Buddhist concepts as suffering, interdependence, and the middle way in responsively reflecting on the possible moral dilemma of AI ethics. The Buddhist middle way, which avoids the extremes of theism and materialism, provides a helpful model for contemplating various thorny issues faced by contemporary managers and policymakers in relation to the latest technological developments.

2 | CAN KILLER ROBOTS BE HELD ACCOUNTABLE?: ROBOTS LACK CONSCIOUSNESS, MIND, AND CONSCIENCE

As an application of AI technology, robots are machines rather than organisms, and even simulations of the human mind lack perception and proprioception, both of which are necessary to truly feel and experience suffering. Suffering and the ending of suffering are the central concern of Buddhism. Since robots do not experience suffering, they lack volition and moral autonomy as karmic subjects, and hence moral regulation would be inapplicable to them.

2.1 | Robots do not experience suffering

Suffering is a recurring theme in Buddhism; Buddhist teachings deal extensively with the topic of cessation of suffering. More than an abstract idea or a metaphysical concept, suffering is concrete experience anchored in body and mind. In Buddhist teachings, it is said that the experiences of sentient beings are founded on the Five Aggregates (Pali *pañca khandha*, Skt. *pañcaskandha*; see Hamilton (1996)): 1. form/body (Skt. *rūpa*), 2. Sensation or feeling (Skt. *Vedanā*), 3. perception (Skt. *saṃjñā*), 4. mental activity or formation (Skt. *saṃskāra*), and 5. consciousness (Skt. *vijñāna*).

Some AI systems seem to possess a rudimentary version of the five *skandhas*, but real suffering is generated through the causal efficiency of karma.[3] Buddhist morality thus would be both simple and practical in nature: suffering is central to the human condition; hence the primary concern of Buddhist moral deliberation is around suffering. Indeed, the straightforward way to distinguish right action from wrong action is to consider the harm caused to others. This is evident in the *Āgama Sutra* or the *Saṃyutta Nikāya* (Bodhi, 2000, p. 1797), where the Buddha employs the idea of "a Dhamma exposition applicable to oneself" (Pali *attupanāyiko dhammapariyāyo*) to stress the importance of abiding by precepts. Some examples of these precepts are no killing, no stealing, no adultery, and no lying. This idea is similar to the Golden Rule, which says that we should treat others the same way we would like to be treated. Since we all wish to attain happiness and cease suffering, we should not harm others just as we wish for no harm to be done to us. In the Buddhist axiological framework, therefore, suffering plays a central role in ethical practice, and it is also an experience only sentient beings can have—suffering beings are sentient beings, and sentient beings cannot escape suffering.[4]

In sum, programmed intelligence is difficult to confer conscious experience,[5] and Buddhist teachings such as the five aggregates and "Dhamma expositions applicable to oneself" suggest that competence of consciousness is necessary for the sensation of suffering and for morality to be feasible. Because AI may lack phenomenal consciousness to feel pain and suffering similar to human beings, it would be inappropriate to attribute moral responsibility to such man-made apparatuses.

2.2 | Simulations of mind ≠ mind

AI may well behave like intelligent beings, yet appearing in the performance of cognition does not mean that they possess the cognitive ability or that their mental makeup is on par with that of humans. The determined responses of such applications are mainly a product of the execution of software programs, such that an outward expression of love remains an algorithm simulated on the actual human experience of love, and the simulations of mind are still different from the real operations of human mind.

One of the most famous arguments against the notion of AI minds is arguably John R. Searle's "Chinese Room Argument." Searle (1991, p. 526) suggests that AI programs are purely formal or syntactical and that the programming itself could not be constituted of nor sufficient for the mental or semantic content of actual human minds. He formulates his argument as follows:

1. Programs are formal (syntactical).
2. Minds have contents (semantic contents).
3. Syntax is not identical with nor sufficient by itself for semantics.

Therefore, programs are not sufficient for nor identical with minds; that is, strong AI is false.

In short, the computing activity of AI is merely a syntactical process, and syntax cannot be equated with semantics; hence a computer program lacks the type of mind needed to understand meaning. Even if an AI were to pass a Turing Test to show its responses are indistinguishable from that of a human, such responses are still mechanical computing—no matter how complex the algorithms may be, mental feeling never arises from it.

Although the above argument was presented some 30 years ago, with immense discussions and disagreements, its theoretical intuition is still inspirational. Searle's version of Biological Naturalism further asserts that only biochemical beings possess conscious experience and meaning understanding. Even though AI applications may be powerful and efficient, their operations remain confined to a mechanical circuit much the same as the way in which cars, television sets, and refrigerators operate. The emphasis on mental understanding of meaning is also what a Buddhist-inspired ethical view on AI would suggest.[6] The Buddhist outlook of five aggregates could be considered congruent with Searle's claims that the function of our conscious mind must presume biochemical elements as constituent parts, including the brain as the central organ of the nervous system that regulates the activities of the feeling body. The "brain" of AI however has no or limited physiological features; AI computational processes do not necessarily result in the consciousness required to guide moral behaviors. That is, moral responsibility demands the agent to be conscious of the moral significance of his or her actions. (Levy, 2014, p. vi) Buddhist ethics stands on this position to advocate that behaviors need mind and consciousness to be granted moral status and entrusted moral obligation.

2.3 | Robots lack volition and moral autonomy

Buddhist morality mainly applies to sentient beings with a capacity for conscience. By saying that "volition (*cetanā*) is karma,"[7] the Buddha equates our mental states with karmic qualities, thus effectively positing mentality as essential for Buddhist morality. Our development of moral dispositions, that is, karmic accumulation, depends on mind–body functions, with mind and consciousness being preconditions for both the experience of suffering and using conscience as a guide to determine the right action. As a robot lacks consciousness and conscience, it is difficult to be considered morally autonomous but only moral heteronomy—being dominated by some custodian and obedient or succumbing to the determinants of another autonomous being.

Looking at the etymology of the term "consciousness," we find that it is closely related to "conscience." The Latin word for consciousness is *conscientia*; in French, it is *conscience*, and in Spanish, *conciencia*. All three terms convey the meaning of both consciousness and conscience. In English, the distinction between consciousness and conscience was only made after the seventeenth century (Jorgensen, 2014; Strohm, 2011, pp. 3–4). Being constituted by the Latin roots *con* and *scientia*, both words carry the implication of "mutual knowledge" and convey a sense of "gathering knowledge" or "encompassing knowledge." Damasio (1999, pp. 230-233 & pp. 309-310) further postulates consciousness as a premise for the higher mental function of conscience. Consciousness, he believes, is a necessary but not sufficient condition for conscience. Since conscience arises on the basis of consciousness, consciousness is required to make ethical judgments feasible.

Since a robot is not capable of freely producing conscious intent, from a Buddhist point of view there can be no moral or karmic imprint per se. Consequently, AI ethics is possible only as ethics *about* robots and not ethics *for* robots. It is mainly human beings that suffer, but not robots. Put differently, any "moral deed" exhibited by an artificially intelligent machine depends on the execution of code that a software engineer wrote. AI ethics concerns the external design and programming of regulative codes for a robot rather than any quality inherent to a robot itself. From a research point of view, AI ethics can apply to matters such as the social, economic, cultural, legal, and political conditions which form the backdrop of the development of AI, but it cannot deal with functions of moral agency on the part of AI.[8] While artificial moral agents can exhibit some form of basic ethical sensitivity to tell "right" from "wrong" (Wallach & Allen, 2010), the decision-making process is neither intentional nor volitional, just as digital art only derives from the creativity and intent of a graphic designer.

Within the Buddhist framework, as mentioned, morality is particularly relevant in matters where suffering and happiness are concerned. Without conscious experience, suffering is not possible. So there would be three key factors which serve as the basis for Buddhist ethics: suffering, mentality (e.g., feeling/*vedanā*, consciousness/*vijñāna*, volition/*cetanā*), and karma. Karma forms from the mental quality of a subject, which in turn relies on his or her deed. Moral decision-making and action do not only pertain to minimizing the suffering of those we interact with; it can also limit our own suffering and increase our own happiness in this life and beyond. Buddhist ethics, therefore, centers around moral cultivation aimed at both mental and karmic purification. The purer our mind, the more regulated our actions, the less potential there is for us to experience suffering.

2.4 | Moral terms are inapplicable to robots

When we say someone is a good person, the word "good" usually carries some degree of moral implication. When we say an AI robot is good, the word "good" instead will refer to the quality of innovation of the human agent who engineered the machine. Moral terms such as responsibility, right, and duty mainly apply to human subjects. They rarely apply to non-human beings. We are difficult to use terms such as compassionate, benevolent, generous, or courageous to praise an AI application, or difficult to denounce it by calling it evil, greedy, stingy, or selfish. When such terms are used on non-human subjects, it would be a form of

personification. Even so, "moral" personifications of non-human subjects do not imply that they are capable of being virtuous but may demand our being kind to them. Rather than talking about "robot rights," it is more meaningful to put emphasis on the duties of human beings and accordingly consider the bestowment of rights to robots a human duty.

According to Buddhist theory, there are six types of beings, that is, *devas*, *asuras*, humans, hell beings, animals, and ghosts, in cyclic existence. Among these, a human rebirth is considered the most fortunate because it affords the greatest opportunity for spiritual growth. As Master Yinshun Shi (1998, pp. 70–71) points out, a human rebirth is the pivot of ascent and descent through the various realms of *samsara*. To be reborn into a satisfactory realm, a being needs to have accumulated positive karma as a human in previous lives. By the same token, misdeeds as a human can lead to rebirth into a miserable realm. Other sentient beings, such as animals, mainly act from instinct. Since their actions lack volition, hunting prey in the animal realm carries only minor karmic retribution for the predator compared to killing in the human realm. Humans are not only capable of devising moral regulations, decrees, and institutions, we also can base our choice on such regulatory frameworks to determine right from wrong. For this reason, humans can be considered the foremost agents of morality.

It seems to me that the Buddhist concept of karma takes three particular factors into account: first, the agent's capacity for moral reflection; second, the target at whom the action is directed; and last, the intention of the agent. Depending on the agent's capacity for moral reflection, wrongful deeds can have greater or lesser consequences. An example may be the applicability of "adultery" or "incest" to inbreeding between animals. Seeing that humans have greater moral awareness, the karmic effects of their actions are greater than those of animals. Second, depending on the virtues of the being at whom the wrongful act is directed, the karmic consequences of a wrongful act such as killing can be more or less severe. For example, killing a human of noble virtue or a benefactor is worse than killing one of a lesser virtue or a stranger. Five Crimes with Immediate Retribution (Skt. *pañcānantarīya-karma*) recorded in the Buddhist texts are: killing your father, killing your mother, killing a liberator/arhat, harming a Buddha's body, and causing a schism in the Sangha community.[9] The consequences of harming a virtuous being whose kindness you may have benefited from are thus far worse than harming an ordinary being. The retribution of killing these beings is so severe that the effects are said to be immediate. (Nyanatiloka, 1980, p. 13, 30, 386) Lastly, the karmic effect of taking the life of a would-be criminal preparing to kill many others will be less severe than killing someone merely out of anger. Killing with intent also carries more severe karmic consequences in comparison to accidental killing, according to the Buddha's teaching that "volition is karma."

So, is the Buddhist idea of karma applicable to AI? Since a robot itself has little to no capacity for autonomous moral reflection and lacks true intentionality, the answer is negative. Even in cases of advanced machine design, the framework that allows for machine learning relies on human engineering. Thus the human designers and users are the ones who are subject to the karmic theory. The applicability of karma to humans suggests we are responsible for developing AI with virtue and wisdom. It implies a robot could potentially be employed as a means to benefit humanity, though it may not be treated as an end. As we enter the Fourth Industrial Revolution, it is vital that we reflect on how to avoid the harmful effects of AI and how to ensure that AI benefits humanity as a whole.

3 | AI—INVENTED BY HUMANS FOR THE BETTERMENT OF HUMANKIND

The Buddhist doctrines of the five *skandhas* and suffering reject a mechanistic position that tends to ignore the organismic and sentient dimensions of the human being. As moral terms are inapplicable directly to robots due to their lack of feeling, human beings are the major agents of morality and virtue. Robotic automation still relies on human autonomy. But the nobility of humankind does not give us the right to dominate other beings. Instead, it should awaken our relational connection with them and entrusts us with obligations to them to empower humanity.

3.1 | The uniqueness of being human

In Buddhism, there are two types of beings: sentient (Pali *satta*, Skr. *sattva*) and non-sentient. Beings with limited or no consciousness, such as plants and minerals, are considered non-sentient. Sentient beings, on the other hand, have the capacity to experience happiness and suffering and continuously come into being, that is, *samsara*, based on their karma. Rebirth as a heavenly being or a human is considered fortunate because less suffering is experienced in these realms compared to the realms of animals, hungry ghosts, or hell-beings. Rebirth as a heavenly being is considered less fortunate than a human being since the level of comfort experienced in the heavenly realms leaves its inhabitants less inclined to engage in spiritual development due to little suffering. Much enjoyment and little trial weaken the motivation to seek liberation. *Homo sapiens* are unique in that they can perceive the living predicaments and have facilities such as rationality and self-awareness to perform righteous deeds and engage in spiritual practices such as meditative concentration and wisdom that lead to enlightenment and cessation of suffering. For this reason, the human condition is both unique and auspicious.[10]

Many Buddhist scripture references emphasize the preciousness of a human rebirth. In the *Saṃyukta Āgama* (CEBTA, T02, no. 99, p. 108) or *Saṃyutta Nikāya* (Bodhi, 2000, pp. 1871–1872), for instance, the Buddha compares the rarity of a human rebirth to the likelihood of a blind turtle surfacing in a vast ocean, at just the right time, and in just the right location, so as to put its head through a small hole in a rotten plank of wood that happens to be floating by. Seen in this light, the rare opportunity that comes with human status is something to be cherished, and the human intellect will never fade

into insignificance compared to the rapid advances of artificially intelligent machines.

In sum, human beings are not only conscious beings but rationally conscious beings. Rationality is also one noteworthy feature of humankind. AI is a sophisticated form of automation that might be intelligent but not consciously rational. The intrinsic characteristic of human "value rationality," as opposed to instrumental rationality, is that it allows us to own up to our moral duties and to face the complexities of modern-day reality. Having the unique qualities does not therefore mean that we should be chauvinistic and elevate ourselves above robots. The Buddhist approach is both humanistic and anthropocentric, yet it does not imply arbitrary domination of other beings. Instead, it emphasizes our obligation to all other forms of being. As humans, we are in the same community and share the same destiny with other forms of being; what is beneficial to others is beneficial to us, and vice versa.[11]

3.2 | The interdependence of humans and robots

The way we as humans deal with non-human beings reflects our inner disposition and mental qualities. The theory of Dependent Origination (Pali *paṭiccasamuppāda*; Skr. *pratītyasamutpāda*) in Buddhism shows that all things arise in dependence upon a cluster of casual factors. This implies that beings have no independent essences but are interrelated to some degree. When we treat non-human beings with kindness, we are directly or indirectly rewarded for bettering their condition. It is thus in our own interest to protect the earth's natural resources as well as to treat all its inhabitants with respect. Despite not being moral agents like human beings, robots as "others" would be similar to animals and plants which are moral patients toward whom moral agents have duties.[12] That said, although a robot does not possess an independent moral status, the robot's creator is human; and the robot is thus a moral object toward which its creator or user has a certain duty of proper manipulation. This holds even truer if the robot or AI benefits humanity, for destroying such an application arbitrarily may be as detrimental to society as killing a guide dog or damaging a historical monument.[13]

According to the 14th Dalai (1999, pp. 162–163), the interests of all beings are interconnected and this realization compels us to develop a sense of "universal responsibility." The Buddhist perspective therefore affords a certain degree of ethical consideration to all beings, both human and non-human. The well-being of non-human beings is also significant because it affects the well-being of humans. Instead of elevating humans above other supposedly inferior forms of being, the Buddhist anthropocentric outlook supports universal kindness to all. In this sense, it is consistent with the concept of universal equality based on dependent origination. For humans to flourish, the welfare of humankind cannot be pursued in isolation from their non-human companions and the environment.

Strictly speaking, we have no direct moral obligation toward robots themselves. What we are responsible for is their appropriate use. Our duty toward other beings shapes human dignity and ensures human happiness. As Grey (1993, pp. 463-475) rightly points out, we need to develop an enriched, fortified anthropocentric notion of human interest to replace the dominant short-term, sectional, and self-regarding conception. A harmonious environment and the convenience of AI technology are two crucial elements that contribute to human welfare, and at the same time, a wholesome human condition is a prerequisite for technological convenience and environmental harmony. Once we develop an awareness of the interconnectedness between our welfare and that of "others," our actions with virtue when designing and operating AI will enhance the value of human existence as well as our life quality.

3.3 | Developing AI from humanity

One feature that differentiates Buddhism from Western theistic traditions is its humanistic approach. The Buddha was human, albeit a human being *par excellence*, but he was not "superhuman" and did not focus on divine or mystical revelations. (Rahula, 1974, pp. 1–4) Liberation is the ultimate goal of Buddhist practice, and the Buddha discovered the path to becoming enlightened through rational and empirical investigations. Thus, Buddhist teachings revolve around the human capacity for positive transformation by integrating insights into our daily lives. As Sen (2014) puts it, the continued relevance and appreciation toward Buddhism over the centuries, or indeed the millennia, is because of the comportment of the Buddha's deep and abiding humanity. This very type of humanity can and ought to play a central role in developing AI.

In order to emphasize the importance of common humanity and shared morality, the Dalai (2012, pp. 16-17) uses an analogy of water and tea to illustrate the relation between morality based on secular ethics and ethics within specific religious beliefs. Secular ethics, he explains, is based on shared morality and can be compared to the water that we *need* every day for health and survival. Ethics based on a religious context is more like tea—tea consists mostly of water and some tea leaves. It is also possible to add spices, sugar, or even salt to make it more nutritious, stimulating, or delicious. The primary component, however, remains water. We can live without tea, but we cannot live without water. In the same way, we can be born free of religious affiliation, but our need for common human values such as compassion is indispensable. Common humanity is more fundamental than any particular religious faith; it also aids in realizing the spiritual goal, so Buddhists shall not disregard conventional morality based on common humanity in their pursuit of religious ideals.

Since both morality and humanity serve as grounds for Buddhist spirituality, the Buddhist path of enlightenment is not contradictory to contemporary civic values such as peace, justice, freedom, equality, mercy, and forgiveness, but instead highlights their necessity. The application of AI technology involves common humanity because, as the creators of such innovations, we are endowed with the type of mental ability that allows us to be considerate toward all the possible consequences of their design. When a robot saves human lives, credit is due to the person who designed the software

and hardware and the person who used it properly, not the robot itself. Similarly, when a robot harms human life, the human agent who misuses should be held responsible, for he or she is the truly conscious rational being whose actions led to the unfortunate outcome.[14] So our responsibility for AI's operation is certain, even if the moral status of AI may be in doubt. The fact that non-human forms of being have limited alertness for moral behavior makes it all the more imperative for us as humans to act with moral accountability. Hence, those who apply AI technology should have a moral sensitivity to guarantee that AI benefits humanity. At the same time, how humans engineer and develop AI applications reflects their own humanity and personality.

3.4 | Advancing humanity through AI

A famous Chinese proverb says that "the water that bears the boat is the same that swallows it up." The same argument can apply to AI—the technology has the potential to bring great benefit to humanity. However, it can also bestow great tragedy by pushing us closer to destructive war. Since it is unwise to sacrifice axiological goals for the sake of scientific progress, the sublimation of humanity should always come before instrumentality. Moral dilemmas involving AI are matters that examine human nature and concern our cognition and connection with artificially intelligent forms of being. The purpose and function of AI ethics are thus more to regulate and protect humans than to be prescriptive toward AI itself.

Scientific advancements often impact our moral integrity in significant ways. It is fair to say that one of our greatest challenges in the 21st century is how to strike a balance between scientific progress and the principles of humanity. So long as humans are the only beings capable of searching for meaning and pursuing a life of value, the rapidly changing landscape of science compels us to continuously rethink our role on this planet. What do we want our future to look like? And how do we go about creating such a future? Stressing the importance of bearing our fundamental values, the Dalai (2006) has said that we need to be critically self-alert about what we develop and why so that we may use new knowledge in the most expedient and ethically responsible manner. Scientists have a particular moral responsibility to ensure that science serves the interests of humanity in the best possible way. Science and technology, therefore, share the same end with spirituality in that both seek the betterment of humankind, albeit through different means.[15]

Because of our unique position as humans, as we embrace innovation and develop the high-tech industry, it is crucial that we keep conscience as our basic guiding principle. The development of AI applications requires a simultaneous recognition of the significance of being human and our role on earth. The future progress of AI technology thus necessitates deeper contemplation of human value and greater wholesomeness of human conduct. From a Buddhist point of view, AI can be employed as a tool for "Bodhisattva practice" because it affords us the opportunity to serve others and ourselves in more efficient ways. Consequently, when founded on the guidance of humanity and morality, the advancement of AI technology could not only support the sublimation of humanity but help realize a better world for all.

4 | A MIDDLE-WAY APPROACH TOWARD THE DEVELOPMENT OF AI TECHNOLOGY

Although we cannot claim that the use of AI systems will definitely contribute to reducing suffering in the long run, at least we have to be certain that it will not further increase suffering. For example, do the robot killers as powerful weapons facilitate the realization of world peace, or do they further increase hatred and pain? Clearly, there are both advantages and disadvantages of AI, as is the case with nuclear science, which can be used for medical treatment and clean energy, or for producing weapons of mass destruction. Similarly, the ruthless efficiency with which an AI weapons system destroys one's enemies is likely to further fan the flames of hatred and conflict. The force of AI may also imply more suffering that humanity has to bear. Instead of setting obstacles to the innovation of AI, it would be more appropriate to make ethics and morality the guiding light behind how it is developed and applied.[16]

The Buddhist philosophy of the middle way applied to moral reasoning provides insights which may minimize harm and maximize benefits.[17] In fact, the Buddhist theory of dependent origination or interdependence that suggested the idea of relationality further implies the Buddhist teaching of the Middle Way (Pali: *Majjhimā paṭipadā*; Skt: *Madhyamā pratipada*). In the *Kaccanagotta Sutta* (Bodhi, 2000, p. 544), the Buddha teaches that one with the "right view" (Pali: *sammā diṭṭhi*; Skt: *samyag dṛṣṭi*) understands that all that exists is subject to dependent arising, and abandons the extreme views of existence and nonexistence.[18] That is, the Buddha teaches dependent-arising as a guide to walking on the middle path. After 6 years of extreme ascetic practice, for instance, the Buddha realized that physical torture is as harmful as sensual indulgence. The philosophy of the middle way thus became one of the central teachings of the Buddha. Also, one of the humanistic features of Buddhism is its avoidance of the two extremes between theism and materialism (Lin, 2020). The Mādhyamaka school founded by Nāgārjuna around 150 C.E. further developed this philosophy (the term *mādhyamika* literally means "one who practices the middle-way") and became one of the principal schools of Mahāyāna Buddhism. Actually, the "middle way" was also taught by Confucius and Aristotle. Although the details are different, all propose avoiding extremes. This may reveal the commonality of human wisdom. So, broadly speaking, not only AI ethics, but all possible moral reflections, could follow the Buddhist spirit of the middle way to cope with tough dilemmas.

Adopting a middle-way approach to AI would enable the prudent evaluation of the threats and benefits of its applications and thereby lead to higher profits with less harm. The development and deployment of autonomous weapons for instance can be appropriate in particular circumstances, such as in defensive operations or when targeting non-human entities, because of their capacity to perform

tasks that are hazardous to human beings, enhance efficiency, lessen casualties, and minimize harm to the innocent. Rather than either advocating for or completely prohibiting autonomous weapons, it may be more advantageous to take a middle approach that prioritizes responsible use. This approach involves incorporating fail-safe measures to prevent unintended harm, creating ethical guidelines and regulations, and establishing legal accountability for their use. By conditionally applying autonomous weapons, we can reduce potential risks and ensure they are used in accordance with international laws and human morality. As such, Buddhist middle-way ethics in AI differs from other frameworks but may incorporate aspects of utility-based ethics, deontological ethics, virtue ethics, and care ethics. It involves expanding utility by taking into account the ethical principles of human responsibility. Utility and duty would be mutually dependent on each other; one cannot ignore the significance of duty in utility, and vice versa. And achieving a balance between utility and duty entails cultivating noble character traits like wisdom, compassion, empathy, and courage. In virtue and care ethics, these traits are prioritized as crucial in directing moral decisions,[19] rather than depending solely on abstract ethical theories. Buddhist ethics of the middle-way may thus offer a holistic and mutually profitable solution to take advantage of various ethical theories.[20]

Technologies, including AI, are considered conceptually neutral, and since they do not have a particular standpoint, embracing the middle-way could guarantee that they are employed to pursue human values rather than promoting political or cultural ideologies. The commonality of humanity requires people to adopt a shared set of values that serve as the foundation of basic moral principles, such as justice, beneficence, efficiency, harmlessness, and respectfulness (e.g., for privacy or autonomy). When grappling with possible conflicts between these values, people can turn to Buddhist philosophy for inspiration, which highlights the importance of cultivating an open and fair mindset toward moral differences, rather than adhering to dogmatic and rigid beliefs. So rather than aiming at the extremes of either maximum freedom of individual choice or collective harmonization of social operation through digitally mediated technology, our application of AI should remain impartial, charting a middle course that could enable the equalization of individual freedom and social stability.

Moreover, the middle-way approach of the Buddhist theory of interdependence urges us to find a balance between creativity and normativity, diversity and unity, security and privacy, solipsism and altruism, etc. Although AI development is often focused on the value of creativity and diversity, in a civic society, values such as normativity and unity are equally crucial. As AI technology quickly collects data on our daily lives, how this pervasive system of monitoring increases our security without invading privacy and sacrificing dignity is also a point in question. Besides, as the convenience of our access to AI increases, so does the amount of time we spend interacting with robots at work and at home; but the convenience of working in solitude should not be obtained at the expense of altruism and concern for real life. Actually, AI applications remain focused on human-to-human connections rather than machine-to-machine or human-to-machine connections. AI should serve the function of a network—connecting people in a spirit of mutual benefit and world peace without exacerbating conflict and strife. Only then will the possibilities of "virtual reality" become a reality informed by virtue.

5 | CONCLUSION

As artificial intelligence pushes to the forefront of technological progress in most fields, the ethical questions surrounding its application and the ripple effects that it could have on society become increasingly relevant. We are at a critical juncture in the development of modern society where instances of humans incurring harm from AI robots or humans causing damage to these robots are no longer uncommon. Aside from posing a threat to job security, AI technologies in the military sphere have the potential to cause a great deal of harm. In this article, however, I specifically drew inspiration from the Buddhist philosophy to shed light on the moral implications of using such technologies in modern society.

Robots, for example, killer robots, cannot be held accountable because of their lack of or limited consciousness, volition, conscience, and so forth. Since suffering is the major consideration of Buddhism, the feeling of suffering is the starting point of moral concern and spiritual development. The fact that AI products are not biological sentient beings, and are deficient in the capacity for intentional deliberation, makes it hard to assign them a legitimate moral status. That said, since a robot lacks a conscious mind to experience suffering, and is "conscienceless," it is impossible to entrust moral duty to such a being. Instead, the moral agents beheld to the actions of AI applications and the consequences of releasing such technologies are its human creators and users. Therefore AI ethics are possible only as ethics *about* robots but not *for* robots.

Robots may, in certain instances, be able to act in more exemplary ways than humans (Sullins, 2010), but any such model behavior is dependent on the execution of code. AI programs may enable robots to fight efficiently on the battlefield as they are the ideal subjects capable of ruling out emotional responses entirely. Yet, morality is not something that is merely functionally computed and mechanically executed. What sets humans apart from robots from a Buddhist standpoint is our conscious feeling and capacity to act virtuously or viciously, which accumulates positive or negative karma. Despite the highly dynamic and sophisticated software programs that govern the outward behavior of advanced AI applications, karma is an axiological concept that cannot be completely designed from algorithms in the Buddhist framework. Since there is no autonomous conscientious agent, the actions of such robots could not be considered karmically relevant. Autonomous will does not arise from programming an AI robot to act seemingly freely.

From a Buddhist perspective, we are in a unique position as humans, and our endowments require that we develop AI for the betterment of humankind. Based on humanity, we could employ the emerging technology of AI to approach the possible evolution of "transhumanity" to enhance the depth of our life reflection and

mental practice.[21] AI can thus be used to improve the human condition and upgrade our overall quality of life. As a result, although there may be paths to "superintelligence" and its related dangers, there will be strategies to confront the challenge. We can solve the "AI control problem," the problem of how to build AI systems to aid without harming their creators or users, if our AI applications have ethical goals that are fully compatible with human prosperity and well-being.

The novelty of this paper, as well as a possible inspiration for future discussions of AI ethics, is to apply Buddhist philosophies to prospect the sustainability of AI development. These philosophies include Buddhism's humanistic thought and the insights of interdependence and middle way. The paper shows how, for instance, a middle-way approach to AI could be used to reduce the risk of AI being utilized in harmful ways. On the contrary, AI advances are fine examples of the achievements of instrumental technology that can be used as a driving factor to realize both individual happiness and social harmony. If employed as means to achieve ends that align with the sublime virtues of humanity, the development of AI technology could serve as an exercise of our noble characters and highlight our significant place as humans on this earth.

FUNDING INFORMATION

No.

CONFLICT OF INTEREST STATEMENT

The author declares no conflicts of interest.

DATA AVAILABILITY STATEMENT

Not applicable.

CODE AVAILABILITY

Not applicable.

ETHICS APPROVAL

Not applicable.

CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

ENDNOTES

[1] Artificial intelligence (AI) here is broadly understood as a computer which uses algorithms to act or think like a human in performing complex tasks and solving various problems.

[2] The specific Buddhist perspective I refer to in this paper is that of Early Buddhism, that is,the teachings recorded in the *Āgamas* or *Nikāya*, which can be said to most closely represent the teachings of the historical Buddha. I also employ Mahayana interpretations in some instances. In fact, it is predictable that Buddhist points of view on AI ethics will continue to develop as AI technology spreads; Soraj Hongladarom's *The Ethics of AI and Robotics: A Buddhist Viewpoint* and Peter D. Hershock's *Buddhism and Intelligent Technology: Toward a More Humane Future* are efforts in this development.

[3] Thanks to an anonymous referee for urging this clarification.

[4] The avoidance of suffering is a goal common to all sentient beings, hence Buddhist precepts may be regarded as both intuitive and universal. In fact, it is not unusual to view suffering as an important aspect of morality. For instance, in his advocacy for animal protection, Peter Singer (1993, p. 61) puts forward the principle of "equal consideration of interest." One of his arguments in favor of giving consideration to animals is that they are equally capable of experiencing suffering and pleasure. Since their natural tendency to pursue happiness and avoid suffering is no different from human beings, the experience of happiness and suffering is where the interest lies. Also see Linzey (2009) for further discussions on this topic.

[5] Although the application of Integrated information theory (IIT) has recently been used to build robots that realize "artificial consciousness" computationally (Kanai et al., 2019), how closely this "consciousness" approximates human consciousness remains unclear. Only if this type of "artificial consciousness" in robots is capable of experiencing suffering would it be necessary to view them as moral subjects. But even then, a robot would have a moral status similar to that of an animal, for example, it would have rights but not duties. Thanks to an anonymous referee for this inspiration.

[6] Although physicalism or materialism is largely considered a mainstream theory of mind and consciousness in contemporary academia, Buddhism rejected the views of the *Lokāyatika* which are in some ways similar to the deterministic worldview of scientific materialism. *Lokāyata*, or *Charvaka* as it also became known after the 7th century, is an ancient Indian school of thought based on physicalism. In the Hindu tradition, adherents of this school propounded both nihilism and atheism, emphasizing direct perception to be the reliable source of knowledge. They were therefore skeptical about the truthfulness of axiology such as karma and moral deeds.

[7] To quote the Buddha, *"It is volition, bhikkhus, that I call kamma. For having willed, one acts by body, speech, or mind."* (AIII, 415; Bodhi, 2012, p. 963) So karma refers to action driven by intention or volition (*cetanā*), which is not confined to the present but may last a long time. Though difficult for non-Buddhists, we more or less presume there is karmic function in our life (or lives) when we accept that there is moral deed or "morality." What is dubious would be the idea of rebirth. Dale Wright (2005) therefore advocates a naturalized concept of karma, that is, a concept of karma disconnected from the assumption of rebirth, and regards the karma as a promising concept in ethical discussions. Also see Harvey (2000, pp. 11–31) for an introduction to the shared foundations of Buddhist ethics as well as the Buddhist theory of karma.

[8] As Boddington (2017, xiv) says, "in many respects, the challenge of AI relates to some of the central presuppositions about the nature of human agents and our place in the world, presuppositions that lie at the heart of ethics itself."

[9] Peter Harvey (2000, p. 138, 413) calls these "five deadly actions" or "five heinous acts." Also see Harvey (2000, pp. 52–58) for more discussions on the various degrees of unwholesomeness in actions.

[10] Although humans are in a better position to pursue axiological lives than any other form of beings, the decline of human morals has the potential to bring about more serious harm. Ignorance, for instance, can drive humans to kill their own children or parents. As the famous Chinese idiom: "even a cruel tiger will not devour its own cubs", humans are worse than animals when they become "animals." See Shi (1971) for more about the "humanistic" features of Buddhism and the significance of being human.

[11] We may not have a moral duty to serve non-human beings such as animals, but acting in an ethically responsible manner does require that we refrain from harming such beings. In his essay entitled "We have only indirect duty to animals," Immanuel Kant suggested that animals are not rational, self-conscious beings capable of grasping laws, hence we do not have direct duties toward them. (Pojman, 2000, pp. 398–399) Kant makes a valid point in suggesting that our moral duty toward non-humans is proportionate to their level of mental awareness and capacity for rational thinking. This does not however mean that we should treat other forms of being without due consideration, since to Kant indirect duties are still duties. Kindness toward animals can for example be helpful in nurturing good character and becoming better accustomed to treating others with compassion.

[12] See Regan (1985, pp. 151–156) for the distinction between a moral agent and a moral patient.

[13] One may wonder which is morally more reproachable: killing an animal or destroying an AI robot? Whether or not it was an endangered species and precisely how its life was ended may influence how we judge the killing of an animal. Also, we seem to have somehow become desensitized to the slaughtering of cows, pigs, sheep, and chickens by virtue of it being a worldwide daily occurrence. Destroying an intricately designed robot almost seems more blameworthy, especially when considering the market price of such a sophisticated piece of machinery versus that of an animal. Certainly, from a "pro-life" point of view, the life of an animal appears to be more valuable than the expedience of a robot. It may be difficult to compare the two, but most would agree that any reckless acts toward them are morally unacceptable, albeit for different reasons.

[14] The inventor of the AI application is its first guardian, and when the guardian hands the application over to someone else, the new keeper or user becomes the virtual guardian and so on. We normally attribute limited responsibility to young children who behave improperly, since they have not yet fully acquired the capacity to distinguish right from wrong. Parents instead are mostly held liable for the actions of their children, because they are the ones accountable for instilling moral values in their offspring and ensuring that they learn to tell the right-wrong distinction.

[15] As the Dalai Lama puts it, since the emergence of modern science, humanity has lived through an engagement between spirituality and science as the two most important sources of knowledge and well-being. Today science and spirituality have the potential to be closer than ever and to embark upon a collaborative endeavor that has far-reaching potential to help humanity meet the challenges before us. See the Dalai (2006, pp. 188–209) for more about his reflections on science and ethics, including the relation between science, spirituality, and humanity.

[16] While there are many substantial rewards of AI, there are various potential risks. If the innovation of AI severely damages relationships between humans and others, either sentient or non-sentient, it cannot be seen as sustainable, since the risks overweigh the benefits. Thus the development and application of AI is not intrinsically good; even though AI is a powerful technology, this in no way ensures that it will be used in a course which is morally correct.

[17] See Hershock (2021, 197–199), who advocates a digital Middle Path as a strategy for resisting and redirecting digital asceticism and digital hedonism, and also Hongladarom (2020, 188), who sees compassion (*karuṇā*) as a middle way between the two extremes of techno-chauvinism and techno-phobism. I am grateful to an anonymous referee for urging a discussion on this point.

[18] As the Kaccanagotta Sutta says,"*But for one who sees the origin of the world as it really is with correct wisdom, there is no notion of nonexistence in regard to the world. And for one who sees the cessation of the world as it really is with correct wisdom, there is no notion of existence in regard to the world*." This sutta reveals the core teaching of the Buddha and is cited in many Mahāyāna texts.

[19] Although care ethics and virtue ethics are not the same, some scholars, for example, Halwani (2003), argue that care ethics should be subsumed under virtue ethics by regarding care as an important virtue.

[20] Thanks to an anonymous referee for this inspiration.

[21] Some suggest that the prosperity of AI technology could assist in going beyond human limitations to envision the possibility of realizing human potential by overcoming aging, sickness, suffering, cognitive limitations, or even death, being optimistic about the development of transhumanism. (Schneider, 2019) But here my usage of "transhumanity" is to imply advancing humanity through using AI to enrich our various capabilities such as morality, stability, compassion, and wisdom, as practices on the way to Buddhahood.

REFERENCES

Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*. Springer International Publishing.

Bodhi, B. (2000). *The connected discourses of the Buddha: A new translation of the Saṃyutta Nikāya*. Wisdom Publications.

Bodhi, B. (2012). *The numerical discourses of the Buddha: A new translation of the Aṅguttara Nikāya*. Wisdom Publications.

Bromwich, J. E. (2019). Why do we hurt robots? *The New York Times* https://www.nytimes.com/2019/01/19/style/why-do-people-hurt-robots.html

Dalai, L. (1999). *Ethics for the New Millennium*. Riverhead Books.

Dalai, L. (2006). *The universe in a single atom: The convergence of science and spirituality*. Three River Press.

Dalai, L. (2012). *Beyond religion: Ethics for a whole world*. Mariner Books.

Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Mifflin Harcourt.

Grey, W. (1993). Anthropocentrism and deep ecology. *Australasian Journal of Philosophy*, *71*(4), 463–475.

Halwani, R. (2003). Care ethics and virtue ethics. *Hypatia*, *18*(3), 161–192.

Hamilton, S. (1996). *Identity and experience: The constitution of the human being according to early Buddhism*. Luzac Oriental.

Harvey, P. (2000). *An introduction to Buddhist ethics: Foundations, values, and issues*. Cambridge University Press.

Hershock, P. D. (2021). *Buddhism and intelligent technology: Toward a more humane future*. Bloomsbury Academic.

Hongladarom, S. (2020). *The ethics of AI and robotics: A Buddhist viewpoint*. Lexington Books.

Jorgensen, L. M. (2014). Seventeenth-century theories of consciousness. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Stanford University. http://plato.stanford.edu/entries/consciousness-17th/

Kanai, R., Fujisawa, I., Tamai, S., Magata, A., & Yasumoto, M. (2019). *Artificial consciousness as a platform for artificial general intelligence*. https://doi.org/10.31219/osf.io/e4jh2

Keown, D. V., Prebish, C. S., & Husted, W. R. (Eds.). (1998). *Buddhism and human rights*. Curzon.

Kessel, J. M. (2019). Killer robots aren't regulated. Yet. *The New York Times* https://www.nytimes.com/2019/12/13/technology/autonomous-weapons-video.html

Krishnan, A. (2009). *Killer robots: Legality and ethicality of autonomous weapons*. Routledge.

Levy, N. (2014). *Consciousness and moral responsibility*. Oxford University Press.

Lin, C.-T. (2020). Could the buddha have been a naturalist? *Sophia*, *59*, 437–456.

Lin, P., & Abney, K. (Eds.). (2014). *Robot ethics: The ethical and social implications of robotics*. The MIT Press.

Linzey, A. (2009). *Why animal suffering matters: Philosophy, theology, and practical ethics*. Oxford University Press.

Nyanatiloka, T. (1980). *Buddhist dictionary: Manual of Buddhist terms and doctrines*. Buddhist Publication Society.
Pojman, L. P. (2000). *Life and death: A reader in moral problems*. Wadsworth Pub.
Rahula, W. (1974). *What the buddha taught*. Grove Press.
Regan, T. (1985). *The case for animal rights*. University of California.
Schneider, S. (2019). *Artificial you: AI and the future of your mind*. Princeton University Press.
Searle, J. R. (1991). Yin and Yang strike out. In D. M. Rosenthal (Ed.), *The nature of mind*. Oxford University Press.
Sen, A. (2014). The contemporary relevance of buddha. *Ethics & International Affairs*, *28*(1), 15–27.
Shi, Y.-S. (1971). *Fozai Renjian* 《佛在人間》*(Buddha in the human realm)*. Zheng Wen (正聞) Publication.
Shi, Y.-S. . (1998). *The way to Buddhahood: Instructions from a modern Chinese master*. Wisdom Publication.Yeung, W. H. trans.
Singer, P. (1993). *Practical ethics*. Cambridge University.
Strohm, P. (2011). *Conscience: A very short introduction*. Oxford University Press.
Sullins, J. P. (2010). RoboWarfare: Can robots be more ethical than humans on the battlefield? *Ethics and Information Technology*, *12*(3), 263–275.
Tzafestas, S. G. (2016). *Roboethics: A navigating overview*. Springer International Publishing.
Waldau, P. (2018). Buddhism and animal rights. In D. Cozort & J. M. Shields (Eds.), *The Oxford handbook of Buddhist ethics* (pp. 650–673). Oxford University Press.
Wallach, W., & Allen, C. (2010). *Moral machines: Teaching robots right from wrong*. Oxford University Press.
Wright, D. (2005). Critical questions towards a naturalized concept of karma in Buddhism. *Journal of Buddhist Ethics*, *12*, 78–93.

AUTHOR BIOGRAPHY

Chien-Te Lin, also known as Kent Lin, is a distinguished professor and director of the Institute of Religion and Humanities at the Buddhist Tzu-Chi University, Taiwan. His main research interests lie in Buddhist philosophy, contemporary Buddhism in Taiwan, Chinese philosophy, and comparative religion (Buddhism and Taoism). His most recent books published in Chinese include *Tao and Emptiness: A Philosophical Dialogue between Laozi and Nāgārjuna* (Dharma Drum Publications, 2013); and *Mind, Consciousness and Liberation: A Comparative Look at Buddhist Theories of Mind and Consciousness* (National Taiwan University Press, 2020). He has also published some English articles in journals such as *Ethical Perspectives*, *Philosophia*, *Contemporary Buddhism*, *Asian Philosophy*, *Sophia*, etc.

How to cite this article: Lin, C.-T. (2023). All about the human: A Buddhist take on AI ethics. *Business Ethics, the Environment & Responsibility*, *00*, 1–10. https://doi.org/10.1111/beer.12547