

The Guiding Role of Buddhist Thought in the Ethics of Artificial Intelligence

Youxin Yan *

Affiliated High School of South China Normal University, Guangzhou, China

* Corresponding Author Email: yanyx.natalie2023@gdhfi.com

Abstract. Artificial intelligence technology has been widely adopted globally, and its application in different fields has brought higher efficiency and accuracy to people around the world, such as the recognition of simple diseases in healthcare, the automatic adaptation of student abilities in education systems, and the planning and construction of smart cities. However, it cannot be ignored that while this technology has brought help to humanity, it has also brought unprecedented machine ethical challenges to human society – such as amplified skin color bias, easy access to personal information, privacy breaches, and the synthesis of false information or images. This paper starts from several core ideas of Buddhism – compassion, meditation, and the interconnection of all things – to analyze the potential guiding role of Buddhist thought in AI ethical issues and discuss exemplary cases of successfully applying Buddhist thought in technological development, thus revealing the important role of Buddhist thought in the future development of AI ethics.

Keywords: Artificial intelligence, Buddhism, ethics.

1. Introduction

In the 21st century, artificial intelligence (AI) has taken root in various fields. AI technology enables machines to perform tasks that traditionally require human skills, such as recognizing and understanding images and text, analyzing and summarizing data, and providing insights based on big data. On the one hand, AI's ability to process and analyze vast amounts of data in a short period of time has greatly impacted various aspects of people's lives. For example, AI assistance in medical procedure has improved the accuracy of disease diagnosis, the application of AI in work has made data processing more convenient, and AI chatbots have helped alleviate people's emotional loneliness.

However, while bringing many positive impacts, the rise of artificial intelligence also presents certain ethical risks. Since it can easily obtain huge amounts of data, biases become prevalent – based on the original unreliability of certain data and the bias contained in the data itself. Often, it is difficult to assess whether the technology is being used to discriminate against others through these biases or is being used to monitor others. Meanwhile, AI technology can be misused in certain cases to create and spread false information, such as the widely circulated fake video of Ukrainian President Volodymyr Zelensky allegedly urging soldiers to lay down their weapons during fierce conflict, without the technology itself being “neutral” to its use.

Facing these problems, some scholars have realized that Buddhist ethical principles could play a constructive and guiding role in reducing the ethical gray areas of AI technology. For example, the Buddhist concept of compassion – a selfless idea of helping others get rid of suffering and gain happiness – means that ethical AI technology should aim to help users reduce suffering and enhance well-being. Moreover, the concept of interconnectedness in Buddhism, which emphasizes the interconnectedness and unity of all things, could effectively guide AI technology to focus on the overall interests of all groups in its development and improvement, rather than providing help to specific groups at the expenses of other groups.

In this era when the formulation of laws and moral codes is far behind the development of technology, it is an urgent issue to formulate moral codes for AI and the high-tech technology to limit their misuse and ensure responsible development. This paper reveals that certain principles in Buddhism can provide corresponding guidelines for its ethics and the moral direction for the

development of AI technologies in the future. Therefore, this paper further explains in detail how specific values in Buddhism can inform and guide the ethical development of AI.

2. Development and Challenges in AI

2.1. Overview of AI Technologies

According to Allen et al., there are generally two types of artificial intelligence: Handcrafted Knowledge AI and Machine Learning AI [1]. Among them, Handcrafted Knowledge AI emerged earlier, while machine learning AI came later. In essence, Handcrafted Knowledge AI resembles to a software development that turns the current knowledge people possess into simple executable programs, that is, programs that accept an “input A” and gives “output b”. Although it is a simple program, many programs in the functioning of society need its help - such as calculating taxes or robot chess players.

Machine Learning AI, by contrast, is what is commonly known as artificial intelligence in modern times - it has learned algorithms and computational models and is able to conduct tasks automatically [1]. This type of AI technology can greatly improve the efficiency of task-performing. In healthcare, Machine Learning AI systems can predict outbreaks of large infectious diseases (such as COVID-19) using predictive models based on previous cases, automatically assess the accuracy of statistical data collected, remove pieces of information that do not reach the reliability standard, and be responsible for some first-step diagnosis of common diseases [2].

In the field of education, platforms embedded with AI technology can perform educational duties independently or together with teachers: on the administrative side, Machine Learning AI systems such as Grammarly and Turnitin can effectively check students’ work (and detect plagiarism and repetition), grade, and finally provide constructive suggestions for further improvement; at the same time, certain Machine Learning AI with adaptive functions can adapt to the learning speed and difficulty according to the learners’ level of comprehension and degree of absorption, which was proved to be more effectively and efficient in the general learning process [3].

In the planning of smart city construction, Machine Learning AI also plays a very important role. For example, the installation of corresponding sensors in garbage cans, transport vehicles and garbage dumps to record and automatically analyze relevant data and make optimization recommendations; corresponding sensors can be installed in the farm to detect the animals and plants’ breeding environment, analyze the multi-dimensional environmental data, and compare the growth state and conditions, so as to customize the best cultivation strategy and minimize the use of pesticides; wireless sensors can be installed in street-side, public and private vehicles to detect pollutant levels and harmful emissions in real time and develop more energy efficient and environmentally friendly solutions based on the recorded data [4]. In conclusion, with this emergence and continuous innovation of AI technology, the way many tasks are performed in society has changed greatly.

2.2. Ethical Challenges in AI

With the increasing popularity and availability of AI technology, some related ethical issues arise synchronously. These issues include, but are not limited to, algorithmic bias, intrusion of personal information, discomfiting monitoring and tracking on personal behaviors and health conditions, and the unpredictable potential for improper use. Algorithmic bias often stems from the bias inherent in the information previously collected by an AI system. Some AI algorithms have been applied to disease diagnosis. However, there is a gap in the accuracy of diagnosis results among different races: black people diagnosed with the same degree of disease by AI are often more seriously ill than white people. This is because in the AI database, black patients previously recorded spent less money on treating the disease and were therefore considered healthier [5].

In addition to the aggravation of bias brought by AI, the potential invasion of personal privacy in the application of AI is also a problem that cannot be ignored. Clearview AI, an AI-related company in the United States, is in the public opinion of privacy violations, which facial recognition system is

widely used by law enforcement agencies in many parts of the United States. It can capture all facial information in any photo ever posted on the Internet, even if individuals in these photos are not aware of being photographed, or if these photos are later deleted by the publisher [6]. Similarly, although the AI technology in smart city construction can make people's lives more convenient, part of the price of this intelligence is the inappropriate surveillance of citizens. To achieve a greater degree of intelligence, citizens' location privacy, footprint privacy is often exploited without their knowledge [7].

A similar AI monitoring scenario can be seen in health interventions: using AI and big data to predict and detect major diseases and make sound policy recommendations that can especially help low-and middle-income countries. However, using AI to supervise patients' disease information still lacks the ethical and moral considerations needed for large-scale deployment [8]. Moreover, in casual shopping places, people's movements and facial information can be tracked without their knowledge: a British chain, the Southern Co-op, has installed cameras that scan customers' faces, which is considered highly unethical, though the store says it is doing so to catch shoplifters and prevent staff from being abused [9].

In addition, because AI technology itself is not able to effectively identify the purpose of its use, this leads to artificial images and videos generated by AI to bring great distress to society. For example, in February 2024, a financial officer of a Hong Kong company was duped into paying \$25 million during a video call – because the fraudsters used synthetic technology to create fake videos of other employees and the company's chief financial officer, leading the financial officer to believe it was real [10]. When evaluating the technology of artificial intelligence, it is obvious that while it brings convenience and progress to contemporary society, the moral and ethical issues it causes cannot be ignored.

2.3. Need for Ethical Frameworks

In modern society, artificial intelligence technology has penetrated all aspects of people's work and life. Some moral problems that have not been properly avoided at present. When using technology, it is important to be aware of the problems of increased prejudice, private information leakage and inappropriate surveillance that it is bringing to individuals and groups. In view of the wide application of AI technology, to avoid existing ethical problems aggravating, the establishment of relevant ethical norms is extremely urgent. In light of the ethical issues raised by AI technology, this code of ethics needs to appropriately advocate respect for all groups, regardless of race, age, gender, emphasize peaceful coexistence among people, and sternly reject living technology to harm others in any way.

3. Buddhist Ethical Principles

3.1. Core Concepts of Buddhism

Buddhism, which originated in India between the 6th and 5th centuries BC, is a religion that emphasizes compassion, selflessness, and interdependence, and is adept at using meditation and mindfulness. Compassion consists of two parts, "co" which means together, and "passion" which means strong feelings. This concept requires that practitioners have a strong capacity for empathy, feeling all emotions with others. Specifically, it refers to the ability to personally feel the pain that others are experiencing, and to sincerely hope that others can escape this pain, willing to take the initiative to take the responsibility to relieve the pain of others. Just as wisdom dominates the rational side of people, compassion influences the emotional side of people [11].

In practice, compassion constantly purifies the motivation of the practitioner. Compassion is like a mirror in which the practitioner constantly gazes at his own reflection, having confirmed that his practice is not to acquire or possess, but to naturally have compassion for himself and others [12]. The method of mindfulness helps practitioners eliminate distractions and perceive things for what they are – by being conscious of what they are doing and focusing their mind on an object. Meditation is a tool for performing mindfulness in Buddhism, where there are four methods of practice to develop

a practitioner's mindfulness skills, including the first step, Kāyānupassanā Satipaṭṭhāna, mindfulness of the body, being aware of one's own body's ongoing movements and feeling how the body works; The second step, Vedānupassanā Satipaṭṭhāna, is to establish mindfulness of feeling, recognizing the feeling you are in at the moment, neutral, pleasant, or unpleasant (note the difference from emotion); The third step, Cittānupassanā Satipaṭṭhāna, is mindfulness of spiritual qualities, such as whether the heart is full of desire, whether there is love; The final step, Dhammānupassanā Satipaṭṭhāna, is mindfulness of mental states, observing the inner, outer, rising and fading of the mind. By carefully observing the continuous moments in which an individual is in countless moments, the practitioner finally becomes aware of the nature of the impermanence of all things (the Buddhist term *anicca*) and becomes aware of the interdependence and conditioning of things [13].

The Buddhist concept of interdependence holds that anything can happen in the world, and that everything depends on people and how things are connected. Everyone, everywhere, and everything depends on other people and things to exist. Every choice one makes throughout one's lives is a significant factor in the direct or indirect consequences of actions. If an individual takes positive actions, the results will be positive, verse vice. Even if anything is possible, the decisions one makes still matter, because these actions affect far more than individuals. In this interdependent environment, an individual's behavior or decision can directly affect other individuals or groups around him, so in such a system, everyone should actively assume their own responsibilities and be responsible for each other [14].

In short, compassion, mindfulness, and interconnectedness are all elements of Buddhist awareness that are necessary to achieve happiness and escape from the suffering of the world. In a strong religious context, these ideas teach people to constantly examine their own actions and their consequences and to look out for the good of others.

3.2. Relevance to Modern Ethics

Buddhist thought not only has a guiding effect on the practice of Buddhist believers, but also shows a kind of good and altruistic propositions in the real society. Compassion is expressed in society as the ability to empathize with others and to restrain oneself from harming others for the benefit of money and power. With compassion, people will naturally pay attention to the vulnerable groups in society, empathize with the harm they have received, feel sorry for it in their hearts, and make actions to help them to finally escape from the harm. For example, when people gradually realize the harm caused to some transgender people or LGBTQ groups by the inherent roles of binary gender established by society, the public will organize demonstrations, and the personnel of the judicial department will establish relevant protection laws to alleviate the persecution and injustice suffered by these vulnerable groups. Otherwise, when any intention arises in our minds that might harm some group in society, compassion should drive an individual to think about the suffering they would face if one were a member of the group being harmed, and to put aside selfish desires and prioritize the interests of the group that might be harmed. If thieves, murderers and other criminals can have a sense of compassion, the crime rate in society will be greatly reduced [15].

Similarly, if scientists can fully consider the advantages and disadvantages of the technology for all groups in society when developing the technology, they can avoid the negative impact of the technology on society as much as possible. Mindfulness is a form of self-awareness through which people understand their environment, the actions they are making, their breath, their feelings, and their mind. When people learn to feel their own existence and the existence of everything around them, and recognize the nature of things, impetuosity will gradually fade from people, and calm thinking will appear. After the acquisition of mindfulness, people can calmly realize the relationship between money, beauty, and power, and get rid of worldly desires, so as to make decisions that are beneficial to society, the environment, and other human beings. If technology developers and policy makers can practice meditation for a period before starting work and making decisions, it has been proven to reduce emotional decision-making and promote positive and beneficial decisions [16].

Likewise, the concept of interconnectedness has many practical ideological implications in society. In society, interconnectedness can be concrete enough to make each person realize that he can exist entirely dependent on other people and things; no human being alone can create everything needed to survive; the books, the food, the movies, and the air all come from the outside world. In the past century, humans have developed unprecedented and very dangerous technologies - such as loggers who can cut down all the trees in the world, but because of the interdependence of all things, human were aware that they will be deprived of the oxygen they depend on to survive if they do so. In a co-existing society, individuals should not only care deeply about the happiness of others, but also sincerely hope that others' prosperity, to help each other in the society and create a more harmonious and prosperous society. If technologists and decision-makers have a strong sense of interconnection, then they will viscerally create technologies and make corresponding decisions that are conducive to social, national, and global stability.

4. The Intersection of Buddhist Thought and AI Ethics

4.1. Compassionate AI Design

In 2020, Banerjeet proposed in his paper the idea of building an AI with a self-aware, embedded empathy model [17]. Artificial intelligence in this model has a self-feedback mechanism - the definition of "having consciousness" here - and is supposed to generate ontological memories and consciousness over time; At the same time, a compassion model will be set up, by analyzing human psychological patterns, so that AI machines can better understand, empathize, and empathize with humans. It can be further speculated that, if they are able to self-examine the underlying moral nature of the instructions they are running, morally conscious AI in the future may be able to autonomously refuse to carry out some instructions that do not meet the moral standards of society, or that would cause unequal harm to certain groups.

While it would be nice to build an idealized AI with absolute compassion, Graves argues that the understanding of suffering and the motivation to respond to it in such a more intelligent AI would be fundamental and difficult to identify [18]. In addition to understanding the Buddhist concept of the universality of suffering, to be truly empathetic, AI needs to be more focused on alleviating the suffering of all beings. At the same time, those responsible for its construction need to ensure that they fully understand the nature of compassion, or else they risk creating machines that do the opposite - that increase the suffering of sentient beings rather than reduce it [18].

4.2. Mindfulness in Design Process

Designers of AI technology should be advised to use mindfulness methods when developing programs and prioritize them in the design process. Firstly, they should learn to listen with empathy and focus on listening to the advice of their partners or superiors before starting work, continuously developing their empathy and empathy skills through this approach. Practicing mindfulness meditation for several times a week can help alleviate anxiety and stress, clear the minds of designers and developers, eliminate distractions, and create higher quality initial ideas and concepts. Learning mindfulness can help them let go of their attachments and expectations, accept all mistakes and accidents that arise in their work, and discover their own problems on this basis, constantly improving. In addition, regular mindfulness can help designers raise awareness of the goal of the program they are developing and supervise the ongoing work to continuously move towards the right path and achieve the result [19]. Artificial intelligence practitioners with this ability can constantly reflect on their previous work achievements and mistakes, adhere to the appropriate ethical standards, constantly review whether the programs they are creating can help society develop better, or whether they will exacerbate unfair biases, stimulate unnecessary conflicts, and make the artificial intelligence technology developed by these personnel more in line with social ethical standards, contributing to the prosperity and development of human society.

4.3. Interconnectedness in Ethical Dilemmas

In order to address the many ethical issues faced by current artificial intelligence technology, the idea of the interconnectivity of all things in Buddhism is highly compatible and effective. This idea of interconnectivity is further manifested as a new moral and ethical standard for global common prosperity. In his book, author Hershock first analyzes why it is not possible to establish ethics applicable to artificial intelligence in a more convenient way - by utilizing many existing regional ethics and expanding their application scope globally [20]. However, it is evident that when applied globally, this approach will inevitably amplify the inherent flaws of these regional ethics, thus establishing a new widely used global ethics. The concept of "prosperity ethics" was first proposed. Under this conceptual framework, people's actions are no longer singular, but rather shift from individual to global perspectives, emphasizing the consideration of the human group. This theory further focuses on non-human subjects and extends to all subjects capable of sensory abilities - these designs and considerations are precisely the application of the theory of the interconnection of all things in Buddhism. This kind of morality not only ensures the well-being of humanity, but also pays attention to the harmony between ecosystems and human society. Harshock further introduced complexity considerations into this theoretical model - emphasizing the moral outlook itself rather than the executor's ability to adapt, which also opened prospects for solving structural problems in artificial intelligence.

4.4. Case Studies and Practical Applications

Although these proposed visions and ideas are far from being completed, some institutions or organizations are already striving to improve, develop, and continuously progress in this area. Mindful Techie is one example of this: it aims to integrate Buddhist ideas such as mindfulness, compassion, and the Internet of Things into the consciousness of technology developers, inspire their empathy and compassion, and assist in the development of artificial intelligence related technologies. In this project, Mindful Techie helps technology developers practice meditation, enabling them to make more comprehensive and thoughtful decisions. At the same time, Mindful Techie's seminars encourage employees to put themselves in the shoes of the end users of technology, inspire their compassion, and comprehensively consider the potential impact of the technology being developed, guiding employees to help others from the perspective of alleviating suffering. Furthermore, Mindful Techie has also made its relevant employees aware of the impact of the technology they are developing on a wider range of social groups, ecological environments, and all human and non-human organisms, making these developers aware of global unity and driving technological output that can promote social progress and development. This institution combines important Buddhist concepts with technological development. Although its influence does not reach a global level, it also reflects the importance that some groups in modern society attach to the ethical challenges of artificial intelligence technology and takes the first few steps towards the integration of Buddhism and AI technology's ethics.

5. Conclusion

This paper mainly introduces the current status of the development of artificial intelligence, the ethical issues it encounters, the main ideas of Buddhism, and how these ideas can theoretically solve the problems currently faced by artificial intelligence to a certain extent. The compassion of Buddhism, mindfulness meditation, and the idea of interconnectivity can theoretically help build a globally applicable new ethical model for artificial intelligence and assist technology developers in improving their current awareness and reflective abilities, ultimately helping to establish AI technologies that reduce bias, alleviate suffering for others, and promote global unity.

In order to better apply the Buddhist ideas mentioned in this article to the development of artificial intelligence, policy makers should pay more attention to the theoretical review of the interweaving of Buddhism and AI technology. Researchers should conduct more investigations and experiments in

this field to fill the theoretical or experimental gaps in existing research. Technicians should try to integrate mindfulness into their technology development work, learn to use meditation to improve their efficiency, and ensure that they comply with ethical standards. Overall, the main ideas of Buddhism have a certain guiding role in addressing the ethical issues of current artificial intelligence and provide a feasible theoretical basis for the ethical challenges faced by artificial intelligence technology, which plays an important role in the future development of AI technology.

References

- [1] Allen, G., Shanahan, J., Allen, G. C., & Shanahan, J. N. (2020). Understanding AI Technology. Joint Artificial Intelligence Center. <https://www.ai.mil/docs/Understanding%20AI%20Technology.pdf>.
- [2] Abdel-Basset, M., Chang, V., & Nabeeh, N. A. (2021). An intelligent framework using disruptive technologies for COVID-19 analysis. *Technological Forecasting and Social Change*, 163, 120431.
- [3] L. Chen, P. Chen and Z. Lin (2020), "Artificial Intelligence in Education: A Review," in *IEEE Access*, vol. 8, pp. 75264 - 75278.
- [4] Appio, F. P., Lima, M., & Paroutis, S. (2019). Understanding Smart Cities: Innovation ecosystems, technological advancements, and societal challenges. *Technological Forecasting and Social Change*, 142, 1 - 14.
- [5] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366 (6464), 447 – 453.
- [6] Duffy, C. (2020, May 29). The ACLU sues Clearview AI, calling the tool an 'unprecedented violation' of privacy rights. CNN. Accessed on September 18, 2024, from: <https://edition.cnn.com/2020/05/28/tech/clearview-ai-aclu-lawsuit/index.html>.
- [7] A. Martinez-Balleste, P. A. Perez-martinez and A. Solanas (2013), "The pursuit of citizens' privacy: a privacy-aware smart city is possible," in *IEEE Communications Magazine*, vol. 51, no. 6, pp. 136 - 141.
- [8] Schwalbe, N., & Wahl, B. (2020). Artificial intelligence and the future of global health. *The Lancet*, 395 (10236), 1579 – 1586.
- [9] Symonds, B. T. (2022, July 25). Convenience store spy cameras face legal challenge. BBC, Accessed on September 18, 2024, from: <https://www.bbc.com/news/uk-england-62297546>.
- [10] Chen, H., & Magramo, K. (2024, February 4). Finance worker pays out \$25 million after video call with deepfake 'chief financial officer.' CNN. Accessed on September 18, 2024, from <https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html>.
- [11] Buddhanet (n.d.). A basic Buddhism guide: wisdom and compassion. Accessed on September 18, 2024, from: <https://www.buddhanet.net/e-learning/qanda07.htm>.
- [12] Barre Center for Buddhist Studies. (2024, August 5). The Nature of Compassion - Barre Center for Buddhist Studies. Accessed on September 18, 2024, from: <https://www.buddhistinquiry.org/article/the-nature-of-compassion/>.
- [13] Admin, & Admin. (2023, April 21). Mindfulness according to the Buddha. Buddho.org. https://buddho.org/mindfulness-according-to-the-buddha/?gad_source=1&gclid=Cj0KCQjwiuC2BhDSARIsALOVfBLhhv6mxnROwpdvVlqcgG_iQ7m0tBLcgOj19F5kFti56yJ3X0IK24oaAtSBEALw_wcB.
- [14] Lion's Roar. (2024). Understanding Emptiness & Interdependence. Accessed on September 18, 2024, from: <https://www.lionsroar.com/understanding-emptiness-interdependence/>.
- [15] Empathy training could reduce crime. (n.d.). Phys.org. Accessed on September 18, 2024, from: <https://phys.org/news/2021-08-empathy-crime.html>.
- [16] Liu, S., Liu, Y., & Ni, Y. (2018). A review of mindfulness improves decision making and future prospects. *Psychology*, 09 (02), 229 – 248.
- [17] Banerjee, S. (2020). A framework for designing compassionate and ethical artificial intelligence and artificial consciousness. *Interdisciplinary Description of Complex Systems*, 18 (2-A), 85 – 95.
- [18] Graves, M., Compson, J., Bhojani, A., Olsen, C., & Arnold, T. (2024). Compassionate AI and the alignment problem. *Theology and Science*, 1 – 5.

- [19] The DYL Team (2024). How mindfulness can transform the design process - designing your life. Designing Your Life. Accessed on September 18, 2024, from: <https://designingyour.life/insights/how-mindfulness-can-transform-the-design-process/>.
- [20] Hershock, P. D. (2021). Buddhism and Intelligent Technology: Toward a More Humane Future. London: Bloomsbury Academic.