ISSN 1974-4382

"THE MORALLY NOBLE PERSON IS NOT A TOOL". AN INTRODUCTION TO CHINESE PERSPECTIVES ON AI ETHICS[1]

ANNA DI TORO
UNIVERSITÀ PER STRANIERI DI SIENA

ditoro@unistrasi.it

Citation: Di Toro, Anna (2026), 'The Morally Noble Person is not a Tool'. An Introduction to Chinese Perspectives on AI Ethics", in Ardizzoni, Sabrina, Marta Aurora, Claudia Buffagni, Anna Di Toro, Imsuk Jung and Andrea Scibetta (eds) *Advanced Technologies, Methods and Materials for Human Health and Well-Being: A Transcultural and Interdisciplinary Perspective*, *mediAzioni* 50: A179–A196, 10.60923/issn.1974-4382/24469, ISSN 1974-4382.

Abstract: The debate on AI ethics in China is as intense as in the Western countries. Chinese philosophical and academic production, as well as official documents, propose values inspired by both Chinese and Western philosophical reflections. The paper aims to examine how the three main Chinese philosophical traditions, *i.e.* Chinese Confucianism, Daoism and Buddhism, consider the ethical issues involved in the use and implementation of new frontier technologies. Chinese ethical approach proposes a "people-centered AI for good" and emphasizes a non-anthropocentric intercultural perspective. These views present many original issues that could contribute to the construction of a globally shared value system for AI ethics.

Keywords: AI ethics in China; official documents on AI ethics; global perspectives on AI ethics.

[1] I would like to express my gratitude to prof. Cong Yali (Beijing University), whose paper "Nursing robot and elder person – perspective on care and emotion under Chinese context", delivered during the Conference *Humans and Machines: Sci-Fi, Ethics and How China is Writing a Shared Future World* (Viterbo, April 2025) offered many inspirations to my research; I would also like to thank the anonymous reviewers for their valuable suggestions.
In the present article, when treating Chinese traditional philosophy, according to the mentality of the time, I use the masculine form as a general form, even if strictly speaking the texts did not specify gender. All translations from Chinese or Italian, where not otherwise indicated, are by the author.

1. Introduction. Intercultural foundational values for AI ethics?

1.1. An overview of Chinese official documents on AI ethics

As in the West, the debate on the ethical implications and risks of the current technological revolution is also intense in China.[2] This debate deserves particular attention: it is essential that Western scholars study the multifaceted Asian contribution in the field of AI ethics and initiate a discussion that integrates reflections from both East and West.[3] As stated by Song Bing (2021b: 10):

> First, foundational values should speak to the totality of humanity and other forms of beings or existence, including perhaps even "conscious" machines in the future. This calls for raising the level of discussion above and beyond individuals, civil organizations, and even nation-states. Secondly, the deployment of frontier technologies is highly distributed, and these technologies are often mutually embedded. They have impacted, and will continue to impact, our political, social, economic, and personal lives, often in unexpected ways. […] Therefore, foundational values should be open, inclusive, and adaptive in this era of frontier technologies. Finally, foundational values should be grounded in the notion of Oneness of all beings…

As we shall see, despite the still widespread stereotype of China as an "immobile" civilization, moved only by exogenous forces, the Three Doctrines (*San jiao* 三教: Confucianism, Daoism and Buddhism) have much to say to contemporary audiences.[4] Moreover, in most of the essays written by Chinese scholars on AI ethics, references to Western philosophy are very frequent, whereas among European or North American authors, there is very little interest in Buddhism, Confucianism and in non-Western traditions in general.

Table 1. Selection of Chinese official documents on AI ethics.

1. Guojia zongtizu 2019, *Report on the analysis of ethical risks of AI*
2. CNNAI (2021), *Ethical Norms for New Generation Artificial Intelligence*
3. CAICT (2022), *Artificial Intelligence White Paper*
4. PRC Position Paper (2022), *Position Paper of the People's Republic of China on Strengthening Ethical Governance of Artificial Intelligence*

As an introduction to the discussion, I will present a selection of the most relevant documents related to AI ethics issued by Chinese official agencies, and

[2] By entering the search key *rengong zhineng lunli* 人工智能伦理 (AI ethics) in CNKI (China National Knowledge Infrastructure, the main academic database in China), we can find 151 articles indexed between Jan. and June 2025. See also the *Report on the analysis of ethical risks of AI* (Guojia zongtizu 2019), promoted by government agencies.
[3] A remarkable step in this direction has been made with the volumes *Intelligence and Wisdom. Artificial Intelligence Meets Chinese Philosophers* (Song 2021a) and *Data Ethics: Building Trust. How Digital Technologies Can Serve Humanity* (Stückelberger and Duggal 2023).
[4] On the stereotype of China's immobilty and on the dinamism of Chinese ideas on the future as discussed by the various philosophical traditions, see Crisma (2010).

10.60923/issn.1974-4382/24469

briefly compare them with documents compiled in European, North-American and global contexts in recent years, focusing my attention on "soft laws" related to values and principles, rather than on "hard laws" (norms and regulations).

The first document to discuss AI ethics in the People's Republic of China (hereinafter PRC) was the *Development Plan for AI of the New Era* (*Xin yi dai rengong zhineng fazhan guihua* 新一代人工智能发展规划, 2017), which stressed the necessity of safety, reliability and controlled development of AI systems. In 2019 the National General Group for the standardization of AI (abbr. Guojia zongtizu) issued the *Report on the Analysis of Ethical Risks of AI*, which presents the national and international discussion and proposes the construction of a global ethical framework. The document emphasizes the centrality of the principle of explainability, in order to offer users the possibility to follow the "reasoning process" of AI and to guarantee human control.[5] Many pages are devoted to the development of regulations concerning protection of privacy and personal data in the PRC, as well as to the risks represented by algorithmic biases. After briefly presenting the ethical standards in countries such as the US, UK, Canada, and Japan, the *Report* proposes two basic principles:

- principle of basic benefit for humanity. AI systems should be aimed at *shan* 善 ("goodness", "kindness"; in the text: *xiangshan* 向善, with an implicit reference to Mencian thought, see 2.2); they should contribute to global peace and environmental protection, and any harmful use of AI for weapons should be avoided. Algorithms should be unbiased and guarantee respect of human rights, freedom, and privacy;
- principle of responsibility. This principle includes two other principles: "transparency", consisting of algorithms that guarantee explicability, verifiability, and predictability; and "consistency of authority and responsibility", *i.e.*, clear accountability throughout the entire process. (Guojia zongtizu 2019: 31).

In 2021, the China National New Generation Artificial Intelligence Governance Specialist Committee released the *Ethical Norms for New Generation Artificial Intelligence* (abbr. *Ethical Norms*).[6]

The document proposes six basic ethical requirements for AI systems that should apply throughout their entire life cycle (*Ethical Norms*: 1):

> the advancement of human welfare, the promotion of fairness and justice, the protection of privacy and security, the assurance of controllability and trustworthiness, the strengthening of accountability, and improvements to the cultivation of ethics.

[5] Interestingly enough, the Deepseek-R1 reasoning model responds to this concern. When it appeared in January 10th, 2025, DeepSeek not only openly shared its algorithms, parameters and training models, but was also the first AI chatbox to explain its reasoning process to the users (see Wang 2025).

[6] The Specialist Committee is under the National New Generation Artificial Intelligence Development Plan Promotion Office (国家新一代人工智能发展规划推进办公室).

10.60923/issn.1974-4382/24469

These norms "aim at incorporating ethics and promoting fairness, justice, harmony [*hexie* 和谐], and security" (*ibid.*: 2). AI activities should promote the Advancement of Human Welfare (Principle 1) by being

> people-centered [*yi ren wei ben* 以人为本], abide by shared human values, respect human rights, appeal to fundamental human interests, and comply with national or regional ethics (*ibid.*).

AI systems should also "promote human–computer harmony and friendliness" (*renji hexie youhao* 人机和谐友好), contribute to ecologically sustainable development, and "jointly build a community of common destiny for humanity" (*ibid.*). The document emphasizes that "humans are the ultimately responsible entities"; and stresses the necessity of heightening "awareness of responsibility, and exercis[ing] self-reflection and self-discipline (*zìxǐng zìlǜ* 自省自律) at every link throughout the AI life cycle" (*ibid.*: 3).

Coherently with the idea that AI systems should be people-centered, abide by shared human values and "comply with national or regional ethics", the 2021 document includes many references to Confucian thought, such as the pursuit of "harmony", the practice of "self-reflection" and the idea of "human–computer harmony and friendliness".

The *Artificial Intelligence White Paper* (abbr. *White Paper*) issued by the China Academy of Information and Communications Technology in 2022, refers to the *Recommendation on the Ethics of Artificial Intelligence*, released by UNESCO in 2021, as a virtuous example of a global normative framework for AI ethics and the broadest consensus reached at the governmental level worldwide.

The document observes that, in the international context, the EU "is steadily advancing from ethics to regulation and desires to take the lead in global AI regulation rules" with the *Artificial Intelligence Act* (2021), the first law in the world to systematically regulate AI (*White Paper*: 19). The document then describes the attitude of the PRC, which takes into consideration both soft and hard laws to promote AI governance, stressing globally shared ethical principles.

The *Position Paper of the People's Republic of China on Strengthening Ethical Governance of Artificial Intelligence* (abbr. *Position Paper*) was issued in 2022. In the document, published by the PRC Ministry of Foreign Affairs and addressed to all countries, China expresses its commitment to the construction of a global governance for AI (*Position Paper*: 1):

> China is committed to building a community with a shared future for humankind in the domain of AI, advocating a people-centered [以人为本] approach and the principle of *AI for good* [智能向善]. China finds it important to enhance the understanding of all countries on AI ethics, and ensure that AI is safe, reliable, controllable, and capable of better empowering global sustainable development and enhancing the common well-being of all humankind. (p. 1)

According to the document, governments should pay particular attention to possible "worst-case scenarios" and identify "potential ethical risks that AI

10.60923/issn.1974-4382/24469

technologies may entail" (*ibid.*: 2). The *Position Paper* emphasizes the necessity to respond to different cultural traditions:

> Governments should […] gradually establish AI ethical systems suited to their national conditions, and improve AI ethical governance mechanisms featuring wide participation and cooperative governance. (*ibid.*)

Governments should moreover "attach importance to public education on AI ethics", "guide all sectors of society to comply with AI ethical rules and norms, and raise AI ethical awareness" (*ibid.*: 3). They should also share with all countries the benefits of AI technologies and the discussion on research, rules-making and ethical issues.

In conclusion, the *Position Paper* proposes to the world a Chinese approach to AI ethics, open to diverse cultural visions, with a "people-centered" approach based on the "AI for good" principle, respecting human rights and the environment.

1.2. What cultural perspectives for AI ethics?

If we compare the above-mentioned Chinese documents on AI ethics with those issued in European, North-American and UN contexts, quite different approaches emerge. All the actors propose their values for the whole world, but Western countries sound more assertive: "technologies should be used to reinforce our highest values", affirms the *US Blueprint for an AI Bill of Rights* (OSTP 2022: 3), while the European Commission's (hereinafter EC) *White Paper* (2020) stresses the creation of an AI "based on European rules and values" to be exported across the world.[7] This global ambition may produce some skepticism, as we can read in Song Bing:

> Many international and inter-governmental organizations and scientific communities have launched campaigns to ensure that their declared principles are the ones that will be adopted as the new norms by the global community. The European Union, for example, made clear its determination to export European values across the world in its AI white paper, published in February 2020 (Song 2021b: 1–2).

UNESCO's *Recommendation* (2021), compiled by an expert group of 24 scholars from all continents,[8] is based on "a holistic, comprehensive, multicultural and

[7] If we may consider the EU's global ambition to export its values as a commendable mission, irony sometimes creeps in among European tech communities, recalling Europe's backwardness in the field of AI: "It's like creating a perfect and sophisticated code of traffic laws when you haven't even begun to pave a single road." ("Deepseek: terremoto nel mondo AI" (live by Datapizza, Jan. 30th 2025: https://www.youtube.com/watch?v=PJeltFJu9Sk, visited 14/07/2025)

[8] In detail, the components of the Ad Hoc Expert Group were from: Uruguay, Ghana, Saudi Arabia, United Arabian Emirates, Russia Federation, Mexico, USA, Poland, Argentina, Latvia, Rwanda, Slovenia, Cameroon, Brazil, Egypt, Morocco, South Africa, India, Japan, France, The Netherlands, UK, Republic of Korea, and China (o.l. https://unesdoc.unesco.org/ark:/48223/pf0000372991, visited 14/06/2025).

10.60923/issn.1974-4382/24469

evolving framework of interdependent values"; it considers ethics as a "dynamic basis for the normative evaluation and guidance of AI technologies", rooted in "human dignity, well-being and prevention of harm" and on environmental protection (p. 10). The Preamble expresses

> concerns about neglecting local knowledge, cultural pluralism, value systems and the demands of global fairness to deal with the positive and negative impact of AI technologies (p. 6).

The principle of centrality of human beings in the use of AI technologies is shared by all the documents, but with different meaningful nuances. EC's *Guidelines* (2019) stress a "human-centric approach […] in the service of humanity and the common good" (p. 4), in which "the human being enjoys a unique and inalienable moral status of primacy" (p. 10), while the Chinese *Position Paper* urges a "people-centered" AI, which should follow the principle of *AI for good*, with no mention whatsoever of human primacy.[9] However, EU documents seem to also include some references to other traditions, as in the mention of the respect due to all "living beings", in Chapter 1 of the *Guidelines*, a pluralistic approach (that recalls Buddhist and Daoist teachings) that we find more clearly expressed in UNESCO's *Recommendation*. In the Preamble of EU *Regulation* (2024), we can observe concern about AI systems used to identify or infer emotions that may "vary considerably across cultures and situations".

By reading the Chinese documents, the references to Chinese philosophical tradition are clear. In the next paragraph, we shall see how the Three Doctrines are involved in the contemporary reflection on AI ethics.

2. Chinese Three Doctrines on AI ethics

In his 2019 essay, entitled "The Possible Influence of the Development of AI on Confucian Ethics", Gan Chunsong offers some valuable insights on the challenge represented by frontier technologies to Confucianism, which is compelled to renovate itself.[10]

As we shall see, all the Three Doctrines accept the innovative challenge of the present technological revolution, offering different answers and positions.

2.1. Humans and AI: harmony as a foundational value

[9] Both Chinese and European documents underline the respect of all living beings among ethical values for AI, but I think that the stress on the "primacy of human beings", in the *EU Guidelines*, reflects a contradiction still present in the Euro-centric approach.

[10] The term "Confucianism" is rather misleading, and many scholars prefer "Ruism", from the Chinese *Ru,* used to refer to the tradition initiated by Confucius (see Song 2016). The term is less personalistic and suggests the historical development of the school. In the present article, however, I use the term "Confucianism", more familiar to Western readers. In its long history, Confucianism underwent many reformulations, in response to contemporary challenges. The deepest one is the so-called "Neo-Confucian" reformulation, initiated in the 11th cent.

10.60923/issn.1974-4382/24469

One of the most original passages contained in the Chinese official documents examined above is the one concerning "human–computer harmony and friendliness" (CNNAI 2021). Both the concepts of harmony and friendship are crucial in Confucian teachings.

Li Chenyang highlights that harmony, in Chinese tradition, is not agreement, but "has to be achieved and maintained with creative tension"; it is a

> dynamic, generative process, which seeks to balance and reconcile differences and conflicts through creativity and mutual transformation (Li 2014: 1).[11]

This complex process requires high morality and self-discipline, together with the capacity of sacrificing one's own interests in the name of righteousness or of collective benefit.[12] All these are characteristics of the *junzi* 君子 ("morally noble person"), the ideal model towards which all human beings should tend.[13] Harmony is thus related to *junzi*, who, thanks to the transformative process of education, develops a "respect for differences while recognizing shared destiny" (Song 2021b: 11):

> 君子和而不同，小人同而不和
> The noble man is in harmony but does not conform [to others]. The mean man conforms [to others], but is not in harmony (*Analecta*, *Zi Lu*, 23).

This dialogue, in which harmony (*he* 和) is matched with non-conformity (*bu tong* 不同), illustrates the consistency of the concept of "human–computer harmony" with Confucian teachings.

Inspired by the Three Doctrines, Song Bing proposes harmony (a concept stressed in Confucianism, but shared by all the Chinese philosophical traditions) and compassion (the core of Mahāyāna Buddhism) as fundamental values for AI ethics, values that should be open, inclusive, adaptive, and able to

> speak to the totality of humanity and other forms of being and existence, including perhaps even "conscious" machines in the future (Song 2021b: 9).

Chinese scholars underline that the Chinese concept of harmony, while presenting many parallels with Heraclitus' idea of harmony as "opposites in concert" and "opposing tension" (Li 2014: 27), is fundamentally different from the "static and structured" mathematical concept in Pythagoras and from Plato's

[11] In 2004, the President of PRC Hu Jintao fostered the idea of "harmonious society"; although the term "harmony" is less recurrent in Xi Jinping's more assertive discourse, it is still an important concept in Chinese political discourse.

[12] On the complexity of the Chinese concept of harmony (possessing in Li Chenyang's opinion the following characteristics: heterogeneity, tension, coordination and cooperation, transformation and growth, renewal) and on its difference with Ancient Greek and Western concepts, see Li 2014: 9–10.

[13] 君子喻於義，小人喻於利 ("The noble man knows what is right; the mean man knows gain", *Analecta*, *Li Ren*, 16).

10.60923/issn.1974-4382/24469

rational, "static and pre-set" harmony (*ibid.*: 31-33). Both concepts are defined as "harmony by conformity" by Li Chenyang, while

> … the ancient Chinese concept of harmony is best understood as a comprehensive process of harmonization, as "deep harmony." It encompasses spatial as well as temporal and metaphysical as well as moral and aesthetic dimensions (Li 2014: 34).

In Song Bing's idea, harmony as a fundamental value for AI ethics would "require us to temper our urge to dismiss and denigrate values and practices which are different from our own" (Song 2021b: 12), thus addressing all humanity. This perspective is also shared by Liu Jeeloo, when she proposes Confucian ethics as a possible basis for AI ethics. According to the Confucian definition of harmony, it is possible to preserve in AI and robots the ethical diversity that we observe in humans (Liu 2022: 687).

Harmony requires adaptation, openness, and self-discipline. In order to achieve harmony in the human–machine relationship, both actors should be able to draw inspiration from the ideal of *junzi.* But since "*Junzi bu qi* 君子不器" ("The noble man is not a tool", *Analecta*, *Wei Zheng*, 12), imbued as he is with morality, AI should also be trained by some form of moral instruction. AI programmers "can set up a learning environment […] for the machine to learn to act in an ethically permissible way" (Liu 2022: 675).

In discussing the risks represented both by self-learning machines that may escape human control and by the hidden biases of human controllers, scholars both from East and West support the *human-in-the-loop* perspective, exploring methods by which humans and machines could positively work together.[14]

The Franciscan theologian and AI expert Paolo Benanti proposes a new discipline in ethics: "algorethics":

> For AI any form of automatic or implicit ethics is unthinkable. It is not possible to make ethics emerge from data. […] we ought to develop a new chapter in ethics: algorethics. [… Algorethics implies] to leave a space for man and his world of values by which to judge: this act may be at times precarious and uncertain, but irreplaceable and not substitutable by machine. [This means] keeping man in the decisional process – *keep human in the loop* – in order to humanize the machine (Benanti 2022: 111).

Similarly to the principle of "harmony and friendliness" between humans and machines, Benanti proposes cooperation between AI (defined by the thinker as "*machina sapiens*") and *homo sapiens*, expressed by the equation *homo sapiens* + *machina sapiens* = (*homo* + *machina*) *sapiens* (*ibid.*: 126).

In order to implement ethical values in AI, scholars discuss the possibility of designing AMAs (Artificial Moral Agents), *i.e.* informatic criteria creating an "artificial morality" in AI systems (see Benanti 2022: 127; Wallach and Allen 2008). But to which ethical models should AMAs adhere? As an alternative to the utilitarian model, which focuses on the most beneficial outcomes for society,

[14] See Monarch 2021; Liu 2022; Benanti 2022.

10.60923/issn.1974-4382/24469

and the deontological model, which focuses on rules, Confucian scholars such as Liu Jeeloo (2022) and Zhu Qin *et al.* (2019) propose virtue ethics:

> Identifying ethical decision with maximizing utility is a dangerous model to be applied to ethical robotics. We do not want to have autonomous robots who always act for the greatest social utility without regard to justice, kindness, responsibilities […], human emotions, interpersonal relationships, and social contexts. Instead of the utilitarian model, I will propose to adopt the model of virtue ethics, in particular of Confucian virtue ethics, in the ethical design of social robots (Liu 2022: 676).[15]

Liu Jeeloo, moreover, underlines that the utilitarian model is action-centered, while the virtue model is agent-centered (*ibid.*: 677). This implies that, in designing AI based on virtue ethics, AI effectively becomes an active agent in a relationship that could also contribute to human ethical development.

The Han Confucian scholar Dong Zhongshu 董仲舒 (179–104 BCE) affirms: "There is no greater virtue than harmony" (*De mo da yu he* 德莫大於和).[16] As suggested by Li Chenyang, however, more than a virtue *per se*, harmony should be rather understood as the harmonizing tension that balances all virtues (Li 2014: 15–16). Harmony is the ultimate achievement in the practice of rites: "As for the use of rites, harmony is the most valuable" (*Li zhi yong, he wei gui* 禮之用，和為贵, *Analecta*, *Xue er*, 12). While *li* 禮 ("rites") contribute to the harmonization of both the intra-personal and interpersonal dimensions, when the natural dynamic way of the cosmos is in harmony, all things thrive, as we find in the *Liji* 禮記: "When *yin* and *yang* harmonize, the myriad things get their due" (*Yin yang he, wanwu de* 陰陽和，萬物得).[17]

In conclusion, the proposal of harmony as a core value in the context of virtue ethics for AI, encompasses the ideal of harmonization in the social, relational, interpersonal dimension (including the human–machine relationship), the harmonizing tension (which does not deny conflicts, but seeks to negotiate them) between different cultures and ethical models, and the harmonization between human society and the natural world.

2.2. Humans and AI: humanity, friendliness and the value of compassion

The value at the very centre of Confucianism is *ren* 仁 (sense of "humanity", "humaneness", "human empathy"). *Ren* has only a few definitions in the *Analecta*: it is defined mainly by the context and is related to *xiao* 孝 ("filial piety"), the value that models the entire Confucian system of social relations:

> 君子務本，本立而道生。孝弟也者，其為仁之本與

[15] For a general introduction to the different models proposed for AI ethics, see Gordon and Nyholm 2021.
[16] *Chunqiu fanlu* 春秋繁露 (*Rich Dew of the Spring and Autumns* [*Chronicle*]), chapter *Xun Tian zhi Dao* 循天之道. https://ctext.org/chun-qiu-fan-lu/xun-tian-zhi-dao (visited 14/10/2025).
[17] *Liji* 禮記 (*Book of Rites*), *Jiao te sheng* 郊特牲 (https://ctext.org/liji/jiao-te-sheng/zhs, visited 13/11/2025).

10.60923/issn.1974-4382/24469

> The noble man devotes himself to the root. Once the root is established, Dao manifests itself. Filial piety and fraternal submission – are they not the root of "humanity"? (*Analecta*, *Xue er*, 2).[18]

Ren, in fact, originates from relations, the character being composed of *ren* 亻 (a form for 人, "person") and *er* 二 ("two"): human beings develop their humanity only in relation with others, and through a painstaking process of learning that integrates the biological nature of human beings and their social one.[19]

The Five Basic Confucian relationships (*Wulun* 五伦, between sovereign and official, father and son, husband and wife, older and younger brother, and between friends) are all hierarchical and two-directional. Their glue bonds are *xiao* 孝 ("filial piety") and *ti* (written as 悌 or 弟, "love and respect for one's older brother"), values that imply a dynamism of duties: not only must the son treat his father with love and respect, but the father has the duty to love, raise, and educate his son.[20] As underlined by Zhu Qin *et al.* (2019), the Confucian idea of friendship is based on moral hierarchy: 无友不如己者 ("Do not make friends with persons that are not your equal", *Analecta*, *Xue er*, 8) and "the ultimate goal in Confucianism is to become a good person through reflective learning in social interactions" Zhu *et al.* (2019: 3). As a consequence, contemporary thinkers assume that AI and social robots too could help people to cultivate *ren* and morality (Zhu *et. al* 2019; Cassauwers 2019). Moreover, in Li Chenyang's opinion, as Confucian ethics are based on the doctrine of "graded love", which allows "differentiated degrees of moral attainment", Confucianism "seems to have more room to accommodate and embrace AI beings as moral agents" (Li 2021: 45).

As underlined by Li Chenyang, the doctrine of "graded love" guarantees harmony, in the dynamic consequentiality expressed by Mencius: since the noble man "is affectionate towards his family, he treats people with humanity; [since] he treats people with humanity, he values things" (親親而仁民，仁民而愛物, *Mengzi*, *Jinxin shang*, 45).

As we have seen, one of the principles proposed by Chinese official documents for AI ethics is *xiangshan* 向善 ("inclined to kindness"; Guojia zongtizu 2019; PRC Position Paper 2022). The Three Doctrines endorse the notion of interrelation and interconnection of all beings; for this reason, "good" encompasses humans, animals, and nature. Confucianism also, originally focused on social relations, then later developed the idea of "oneness" of humans, animals and nature. Zhang Zai 张载 (1020–1077), a Confucian scholar of the Song dynasty, affirms: "All people are my brothers and sisters, and all things are my companions" (民，吾同胞。物，吾与也).[21] This vision reflects the Buddhist idea of interrelation of all beings: "in Buddhist thinking, humans, animals, and

[18] Transl. by J. Legge modified by the author.
[19] On the relation between the biological and social nature of man in Confucianism, see Gan 2019.
[20] In spite of the centrality of *xiao* 孝 ("filial piety") in Confucianism, contemporary thinkers prefer not to include it among the values of AI ethics, since it is strictly linked to human family ties and emotions (Liu 2021: 16).
[21] Zhang Zai, "Western Inscription", in W.T. de Bary and I. Bloom (eds) *Sources of Chinese Tradition*, 2nd ed., Vol. 1. New York: Columbia University Press, 1999, 683. Quoted in Song 2021b: 4.

10.60923/issn.1974-4382/24469

nature are all manifestations of Being" (Song 2021b: 5). As for Daoism, "in the light of the Dao, all things are equal" (*ibid.*: 4).

The Three Doctrines all endorse the concept of inter-relation and oneness of all beings, albeit with different nuances. Mencius connects the capacity to "cherish things" (*aiwu* 愛物 – i.e. the material world) to the practice of *ren* 仁 ("humanity") towards people, which is learnt by practising "affection" (*qin* 親) within the family. Love and respect for the material world stem from an individual's moral growth. In Daoism and Buddhism, the oneness of all beings is of ontological nature. In order to develop this concept in AI, along with "harmony", Song Bing proposes "compassion" in the Buddhist sense as a foundational ethical value for AI, which differs from the Christian concept of "pity".

The XIV Dalai Lama thus defines "compassion": "it does not imply pity. […] There is no sense of condescension. On the contrary, compassion denotes a feeling of connection with others, reflecting its origins in empathy."[22] Buddhist "compassion" is pervasive and spontaneous because it is based on the notion of interconnectivity between all beings (far beyond human society) and on the experience of suffering, equally shared by all sentient creatures. As we have seen in the UNESCO *Recommendation on the Ethics of Artificial Intelligence*, Buddhist inspiration is particularly effective in terms of giving AI the value of respect for environment and life in all its forms.

Going back to the *xiangshan* 向善 principle, according to Mencius, *shan* operates a virtuous relational circle:

> 取諸人以為善、是與人為善者也、故君子莫大乎與人為善。
> To take example from others to practice kindness is to help them in the same practice. Therefore, there is no attribute of the superior man greater than his helping men to practice kindness.[23]

An AI imbued with the *xiangshan* principle can act as an educational model. This idea is also expressed in the 7th rule of the "Confucian Robotics Ethics" code proposed by Liu Jeeloo: "A robot must render assistance to other human beings in their pursuit of moral improvement" (Liu 2023: 200). As a consequence, robots and AI could contribute to the social project of human self-cultivation (Zhu *et al.* 2019).

2.3. Humans and AI: learning as a transformative force

At this point, another fundamental part of Confucian teaching comes into play: *xue* 学 ("to study", "to learn"). *S*ince the aim of Confucius is to render human beings really and entirely human (Cheng 2000), this process begins with the decision to learn, as we read in Confucius' ideal autobiography: "At fifteen, I had

[22] Gyatso Tenzin, *Ethics for the new millennium*, New York: Riverhead Books, 1999, quoted in Song 2021b: 10.
[23] *Mengzi* 孟子, "Gong Sun Chou shang"《公孫丑上》(online version by J. Legge available at https://ctext.org/mengzi/gong-sun-chou-i (visited 18/07/2025); the translation has been slightly modified by the author.

10.60923/issn.1974-4382/24469

my mind bent on learning" (吾十有五而志于學, *Analecta*, *Wei zheng*, 4; transl. J. Legge), an action that becomes agreeable in the end: "Is it not pleasant to learn and constantly apply [what you learn]?" (學而時習之，不亦說乎, *Analecta*, *Xue er*, 1).

The centrality of learning is at the core of Chinese philosophy, not only of Confucianism. As underlined by Gan Chunsong (2019), the *Sanhuang wudi* 三皇五帝 (Three Sovereings and Five Emperors) of Chinese mythology taught human beings the basic techniques for survival (fishing, hunting, farming, use of medical herbs, medical treatments, etc.). These forefathers also taught humanity the *bagua* 八卦, "Eight Diagrams", that symbolize natural phenomena and help men interpret nature, using the *Yijing* 易經 (*Classic of Changes*), a divination manual rich in philosophical reflections. Despite having different approaches to nature, all these techniques are not based on subjugating nature, but on creating harmonious relations between nature and human beings.

As observed by Song Bing (2021b: 5), non-anthropocentrism is a feature of Chinese traditional philosophy: "none of the three dominant schools of Chinese thinking places human beings in a supreme and crowning position within the universe". As a consequence, this

> strong non-anthropocentrism within the dominant Chinese philosophical schools has contributed to a relatively open, if not entirely relaxed, attitude towards the rise of the "super-power" of AI and robotics in China in recent years (*ibid.*).

Scholars from all countries reflect on the fact that AI systems seem able to duplicate the unique features of human intelligence, threatening the special status of humanity.[24] In reflecting on this discussion, Li Chenyang observes that AI could help us discover that we are not so special: in the Daoist perspective,

> between humanity and other existing things in the world, there are differences without distinctions. It is entirely possible […] that our future lies in integrating with advanced AI technology rather than maintaining our distinctiveness (Li 2021: 36).

Li theorizes that, if AI shall lead to the end of humanity in its traditional sense, it could also "extend human existence into new territories" (*ibid.*: 37).

Nonetheless, according to the Daoist tradition, new technologies may alter the pure natural simplicity of things, inasmuch as technologies make them deviate from the spontaneous (*ziran* 自然, "spontaneity", "naturalness", "nature") flow of Dao. Robin R. Wang quotes Laozi:

> 人法地，地法天，天法道，道法自然。

[24] On the Western debate of AI's threat to the uniqueness of human beings, see Benanti 2022, Chap. 1.

10.60923/issn.1974-4382/24469

> The law of man is [given by] the Earth, the law of the Earth is [given by] Heaven, the law of Heaven is [given by] the Dao, the law of Dao is spontaneity (*Dao De Jing*, Chap. 25).[25]

Robin Wang observes that *ziran* is "a human's most potent mode of action" and, despite being created by humans, "AI cannot grasp *ziran*", which is grounded on the "mysterious efficacy" (*xuande* 玄德) of the spontaneous and unpredictable movement of the self-generating and inter-related forces of *yin* 阴 and *yang* 阳 (Wang 2021: 75). At the same time, "machines cannot flow like Dao", whose flow "relies on *shen* 神, the spirit" that "can inhabit the human body" (Wang 2021: 71 and 70). However, in Robin Wang's opinion, AI could learn from Daoism to further expand the binary logic which is at the heart of computer science, by developing the peculiar ability that the scholar calls "yinyang intelligence", an intelligence able to "resonate with the hidden forces at work" and, similarly to human intelligence, is "always ready to adapt to the unexpected" (Wang 2021: 77). The thinker leaves the question open for further reflection, linked to the fundamental question of whether AI can possess creativity and generate groundbreaking ideas.[26]

Going back to *xue* ("to learn"), its transformational force could also involve AI agents. However, as underlined by Gan Chunsong (2019), the emotive factor, made of joys and frustrations, is a fundamental part of the learning process. As expressed by the maxim *xue yi cheng ren* 学以成人 ("study in order to become a person"), to learn means to be able to restrain one's natural instincts and desires through the fundamental teacher–pupil relation by which we apprehend the rules of our social existence. For this reason, Gan Chunsong (2019) is particularly cautious when reasoning about the risks of delegating our intellectual activities to AI. The scholar underlines that AI cannot be an emotive intelligence, and all the social function of education is lost when the learning process is entrusted to AI systems.

Education is at the core of the two main ancient interpreters of Confucian thought, Mencius and Xunzi. Xunzi argued that, being humans evil by nature, the only way to transform them is through strict education. In Xunzi's idea, ethics is "socially made or artificial (*wei* 偽)" (Li 2021: 43) and men are artificially transformed into moral and social beings thanks to rigorous rules. Li Chenyang thus proposes the Xunzian perspective as a possible approach to AI ethics, since it "would primarily be about devising effective rules" (Li 2021: 43). AI ethics should also be, in Li Chenyang's opinion, Mencian. In Mencius's teaching, the difference between humans and beasts "lies with the possession of a kind heart", which is bestowed on humans by nature but preserved and developed by education.[27] In order to reach the status of moral agents, AI systems, especially

[25] Translation by the author, based on J. Legge.
[26] The research on the question is still in an early stage. For a Bibliography on the subject, see: https://salve.libguides.com/aicreativity.
[27] We should underline that in Confucian thought education is linked also to the fundamental sphere of *li* 禮 ("ritual"), which has a religious, social and formative function.

10.60923/issn.1974-4382/24469

those used in social and care contexts, should be given a "heart" (*xin* 心, Li 2021: 42) based on the *xiangshan* (AI for good) principle.[28]

In my opinion, however, the idea of endowing AI with *xin* is not free of contradictions. *Xin* (often translated as "heart-mind") in Chinese tradition is both physical and moral, emotive and rational, it is "not only the seat of cognition but also of emotions and desires".[29] According to Chinese philosophy, it is not possible to separate cognition and body.[30] The question is: is it possible to imagine disembodied virtues? Is it possible to imbue empathy in beings with no *pathos* ("feeling")?

3. Concluding remarks

An intelligence with no body and feeling is understandably alarming. But maybe we should shift our reasoning to the influence the relationship with AI beings will exercise on our lives. Ugo Morelli, in his article "Chi ha paura dell"AI?" ("Who is afraid of AI?") observes:

> If AI is at the same time the product and the model by which we explain the functioning of our mind, it is not possible to reduce the comprehension of ourselves to an algorithmic question. We are not just language and cognition, but the dynamic expression of our corporeal and relational nature. "Embodied cognition" is […] an appropriate point of view to study the impact of digital systems on ourselves and on social relations, and inter-corporeity is acknowledged as the primary source of our cognition of others (Morelli 2025: 4).

Fear of AI does not help the development of a healthy attitude towards these technologies. In the West, however, technophobia seems to be mainstream (see Aresu 2024). In reading the reflections of some Chinese thinkers, we may find apparently similar positions:

> The rapid advancement of Artificial Intelligence technology means the end of humanity in an important sense. We will irreversibly lose the special status that we have claimed to possess (Li 2021: 35).

However, Li Chenyang's is not a technophobic attitude: reinforcing the idea that human beings do not have a special status in nature, AI technology could help us discover that

[28] Li Chenyang quotes John Havens' "heartificial intelligence" (J. Havens, *Heartificial Intelligence: Emrbacing our humanity to maximize machines*, Toronto: Tarcher/Penguin, 2016)
[29] See "Mind (Heart-Mind) in Chinese Philosophy" Stanford Encyclopedia of Philosophy (https://plato.stanford.edu/entries/chinese-mind/ (visited 21/08/2025).
[30] We should underline that also Western contemporary thinkers share this perspective: "Our cognition and intelligence are such because they are embodied" (Benanti 2022: 110).

10.60923/issn.1974-4382/24469

> humanity is not so distinctive and learn to live with the consequences of such a discovery. Daoism is perhaps more ready to accept such a conclusion than many other philosophies (*ibid.*: 36).

As we have seen, the "strong non-anthropocentrism" of Chinese thought (see 2.3) is echoed also in the Chinese official documents on AI ethics that promote "human-computer harmony and friendliness" and the idea of AI as a moral educational model (CNNAI 2021).

Both Chinese and Western thinkers underline the importance of preserving the human factor at the centre of human-AI interactions. Paolo Benanti proposes algorethics in order to *keep man in the loop*. In the frame of the "techno-human condition" in which we live, we should carefully (re-)establish the terms of our relation with reality through technology (Benanti 2022: 11). Technological advancement not only enhances our exploitation of and control over nature, but deepens our knowledge of ourselves and requires a continuous ethical and philosophical reflection (Gan 2019).

If in most of the philosophical traditions of ancient China the attitude towards technology was substantially positive, as long as it did not go against nature, only a few schools (such as Daoism and the Yin-yang school; see Needham 1956) developed a scientific thought, which remained marginal. In Confucianism and the entire long *literati* tradition, "there was no room for science, only for traditional technology" (Needham 1956: 29; see also Gan 2019). For this reason, China does not have a tradition of philosophy of science that reflects on the relationship between man and technology. In Gan's opinion, in the face of new contemporary challenges represented in particular by AI systems and biotechnologies, the contribution of philosophy and ethics and an intense communication between scientists and philosophers are urgent needs, since biotechnologies threaten the Confucian-based Chinese family model by challenging blood ties and the efficacy of traditional roles, and AI threatens the learning process and the educative role of family and society.

Gan Chunsong concludes:

> What should philosophers do? They should establish collaborative relations with scientists: on the one hand, they could help scientists comprehend […] how the ancient sources could contribute to the values that define our contemporary meaning of life. On the other hand, they could integrate AI into the possible space of expansion of humanity, by reflecting on how contemporary Chinese people should conceive human nature and human values and […] propose Chinese values such as affection for relatives, family ties, filial piety, etc. for the future development of humanity. Through this interaction, Chinese philosophers and scientists could contribute to the development of Chinese thought with issues that belong to Chinese traditional philosophy itself (Gan 2019: 15).

In conclusion, if AI is destined to become a friendly moral model for humans that follows the Mencian *xiangshan* ("aimed at goodness") principle, and thereby acquire the status of a moral agent, we should also keep in mind, when training AI, that "The noble man (*junzi*) is not a tool" (*Analecta*, *Wei Zheng*, 12). This

10.60923/issn.1974-4382/24469

maxim emphasizes the wholeness of the ethical being, a being who "should possess the universality of the Way".[31] In *junzi* no action can be separated from morality, a sense of justice and the aim of benefitting others, since "The noble man knows what is right; the mean man knows gain" (君子喻於義，小人喻於利, *Analecta*, *Li Ren*, 16). Moreover, *junzi* cannot be used as a tool: *junzi* is an ideal towards which human beings should tend. He/she develops through a complex and laborious learning process involving body and mind, cognition and emotion, during which values are embodied through study and behavior is internalized through the practice of ritual. All these processes contribute to the development of an ethical being capable of living harmoniously with him/herself, others and nature. Is this development possible for Artificial Intelligence?

REFERENCES

Aresu, A. (2024) *Geopolitica dell'intelligenza artificiale*, Milano: Feltrinelli.

Benanti, P. (2022) *Human in the Loop. Decisioni umane e intelligenze artificiali*, Milano: Mondadori.

CAICT (China Academy of Information and Communications Technology) (2022) *Artificial Intelligence White Paper*, Engl. version. Available at https://cset.georgetown.edu/publication/artificial-intelligence-white-paper-2022/ (visited 07/06/2025).

Cassauwers, T. (2019, March 28) *How Confucian Could Put Fears about Artificial Intelligence to Bed*. Available at https://www.ozy.com/immodest-proposal/how-confucianism-could-put-fearsabout-artificial-intelligence-to-bed/93206 (visited 05/07/2025).

Cheng, A. (2000), *Storia del pensiero cinese*, Torino: Einaudi, vol. 1.

CNNAI (China National New Generation Artificial Intelligence Governance Specialist Committee) (2021) *Xinyidai rengong zhineng lunli guifan* 新一代人工智能伦理规范(*Ethical Norms for New Generation Artificial Intelligence*) Engl. version. Available at https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/ (visited 05/06/2025).

Crisma, A. (2010) "Idee di futuro nelle tradizioni di pensiero cinesi", *Giornale Critico di Storia delle Idee* 2(3): 191–206.

EC (European Commission) (2019) *Ethics Guidelines for Trustworthy AI*, November 2019. Available at https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (visited 10/06/2025).

------ (2020) *White Paper on Artificial Intelligence. A European approach to excellence and trust*, February 2020. Available at https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (visited 08/06/2025).

[31] I quote Cong Yali's paper "Nursing Robotics and Moral Relationship Growing", presented in the Conference *Inclusive Perspectives on Care: Mediation, Narrative Medicine, and Assistive Robotics*, Siena, Siena University Hospital and Università per Stranieri, Nov. 6th, 2025.

10.60923/issn.1974-4382/24469

EU (European Union) (2024) *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence*. Available at https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (visited 05/06/2025).

Gan, C. 干春松 (2019) "Rengong zhineng de fazhan dui Rujia lunli suo keneng dailai de yingxiang" 人工智能的发展对儒家伦理所可能带来的影响 (The Possible Influence of the Development of AI on Confucian Ethics), *Kongzi yanjiu* 5: 38–47. Available at https://www.rujiazg.com/article/18627 (visited 05/06/2025).

Gordon J.-S. and S. Nyholm (2021) *Ethics of Artificial Intelligence*, *Internet Encyclopedia of Philosophy*. Available at https://iep.utm.edu/ethics-of-artificial-intelligence/ (visited 05/07/2025).

Guojia zongtizu (Guojia rengong zhineng biaozhunhua zongtizu 国家人工智能标准化总体组) (2019) *Rengong zhineng lunli fengxian fenxi baogao* 人工 智 能伦理风险分析报告 (*Report on the Analysis of Ethical Risks of AI*), April 2019. Available at https://www.cesi.cn/images/editor/20190425/20190425142632634001.pdf (visited 05/06/2025).

Li C. (2021) "The Artificial Intelligence Challenge and the End of Humanity", in B. Song (ed), 33–48.

------ (2014) *The Confucian philosophy of harmony*, London: Routledge.

Liu J. (2022) "Human-in-the-Loop Ethical AI for Care Robots and Confucian Virtue Ethics", in F. Cavallo, J.-J. Cabibihan, L. Fiorini, A. Sorrentino, H. He, X. Liu, Y. Matsumoto and S.S. Ge (eds) *Social Robotics: 14th International Conference ICSR 2022 Proceedings, LNAI 13818*, Part II, Berlin: Springer, 674–688.

------ (2023) "Confucian Robotic Ethics", in C. Stückelberger and P. Duggal (eds) 175–207. Available at https://core.ac.uk/download/582653514.pdf (visited 08/07/2025).

Monarch (Munro) R. (2021) *Human-in-the-Loop Machine Learning. Active learning and annotation for human-centered AI*, New York: Manning

Morelli, U. (2025) "Chi ha paura dell'AI", *Doppiozero*, May 30. Available at https://www.doppiozero.com/chi-ha-paura-dellai (visited 08/06/2025).

Needham, J. (1956) *Science and Civilization in China*, Vol. 2: *History of Scientific Thought*, Cambridge: Cambridge University Press.

OSTP (Office of Science and Technology Policy) (2022) *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*. Available at https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/ (visited 05/07/2025).

PRC Position Paper (2022) 中国关于加强人工智能伦理治理的立场文件 (*Position Paper of the People's Republic of China on Strengthening Ethical Governance of Artificial Intelligence)*, issued by China's Ministry of Foreign Affairs. Available at https://www.mfa.gov.cn/ziliao_674904/zcwj_674915/202211/t20221117_10976728.shtml; Engl.:

https://www.fmprc.gov.cn/eng/zy/wjzc/202405/t20240531_11367525.html (visited 05/07/2025).

Song, B. (2016) “A Cathechism of Confucianism: is Confucius a Confucian?”, *Huffington Post*, Feb. 9. Available at https://www.huffpost.com/entry/a-catechism-of-confuciani_b_9178068 (visited 20/05/2025).

------ (ed) (2021a) *Intelligence and Wisdom. Artificial Intelligence Meets Chinese Philosophers*, Berlin: Springer.

------. (2021b) “Introduction: How Chinese Philosophers Think About Artificial Intelligence?”, in Song, B. (ed), 1–14

Stückelberger C. and P. Duggal (eds) (2023) *Data Ethics: Building Trust. How Digital Technologies Can Serve Humanity*, Geneva: Globethics Publications. Available at https://core.ac.uk/download/582653514.pdf (visited 20/07/2025).

UNESCO (2021; updated 2024) *Recommendation on the Ethics of Artificial Intelligence*. Available at https://unesdoc.unesco.org/ark:/48223/pf0000381137 (visited 15/06/2025).

Wang G. (2025) “DeepSeek Has More to Offer beyond Efficiency: Explainable AI. Available at https://www.forbes.com/sites/geruiwang/2025/01/30/deepseek-redefines-ai-with-explainable-reasoning-and-open-innovation/ (visited 12/07/2025).

Wallach W. and C. Allen (2008) *Moral Machines: Teaching Robots Right from Wrong*, Oxford: Oxford University Press.

Zhu Q., T. Williams and R. Wen (2019) “Confucian Robot Ethics”, in D. Wittkower (ed) *2019 Computer Ethics – Philosophical Enquiry (CEPE) Proceedings*, Vol. 2019, Article 12. Available at https://digitalcommons.odu.edu/cepe_proceedings/vol2019/iss1/12 (visited 21/08/2025).

WEBSITES

CNKI (China National Knowledge Infrastructure): https://www.cnki.net/index/