Stimuli from selected non-Western approaches to AI ethics

Soenke Ziesche
Independent researcher
Delhi, India
soenke.ziesche@gmail.com

Abstract

While the urgent need for ethics for the thriving field of AI has been acknowledged, currently Western approaches to AI ethics are prevalent. This constitutes a problem because, on the one hand, these approaches tend to reflect the values of the regions where they are originating from, on the other hand, not all values are universal. This form of digital neo-colonialism ought to be prevented. As a step in this direction this article presents ten selected concepts of non-Western approaches to AI ethics and analyses their originality as well as potential combability with the Western approaches. Based on this, the article concludes with a recommendation to merge Western and non-Western approaches towards universal AI ethics as far as they are compatible and to attempt to reconcile those aspects, which appear incompatible.

Keywords

AI ethics, non-Western, Buddhism, Confucianism, Hinduism, Islam, Maori, Shintoism, Ubuntu

Introduction

The enormous developments in the field of AI in the past decade urgently necessitate ethics for AI, given that AI is another dual-use technology and a range of risks and dangers of AI has become evident (e.g., Brundage et al., 2018). Albeit with some delay, there are by now a substantial number of approaches to AI ethics (e.g., for overviews: Jobin et al., 2019 or Corrêa et al., 2022). However, the vast majority of them is derived from Western countries. For example, an inventory by Algorithm Watch contained 167 AI ethics guidelines in 2020, of which 15 were merely from Asia and one from Africa.[1]

This constitutes a problem because, on the one hand, guidelines tend to reflect the values of the regions where they are originating from, on the other hand, not all values are universal. In other words, currently AI ethics guidelines are dominated by Western values, while values from other parts of the world are not appropriately represented. AI ethics guidelines cover a

[1] https://inventory.algorithmwatch.org/

range of issues, including which and how ethical principles are implemented in the decision-making of AI systems. These ethical principles should not be based exclusively on Western values. To prevent such a form of digital neo-colonialism, it would be highly desirable if AI ethics guidelines take equitably into account the rich diversity of value systems, traditions and ideas generated in many centuries by the many cultures in the world.

As a caveat from a historic perspective it has to be mentioned that a complete solution for this problem is perhaps not realistic. After all, the issue of unbalanced AI ethics can be seen as a reflection of existing issues in the real world. Throughout human history there have been conflicts, many of which originated due to incompatible value systems. And up to the present day value systems prevail in parts of the world, which are conflicting.

Nevertheless, against this backdrop, a variety of scholarly proposals have been made for non-Western approaches to AI ethics. Such proposals are supported by the landmark UNESCO "Recommendation on the Ethics of Artificial Intelligence", which states that "the objectives of this Recommendation are: (a)to provide a universal framework of values, principles and actions to guide States in the formulation of their legislation, policies or other instruments regarding AI, consistent with international law." (UNESCO, 2021, p.5).

Research objectives

In this article selected proposals for non-Western approaches to AI ethics are reviewed to identify aspects, which have not been considered in Western AI ethics. In a second step, these aspects are analysed according to the research question whether these aspects are compatible with Western AI ethics; thus, these approaches to AI ethics could be merged, or whether these aspects are incompatible with Western AI ethics; thus, these approaches to AI ethics are conflicting.

The outcome of this research can be seen as groundwork as well as prerequisite of improved universal AI ethics guidelines and diversified ethical principles for AI decision making in particular, which represent the values of all people.

It must be noted that this paper is not another inventory of AI ethics, but looks at samples of non-Western approaches to AI ethics. Also, in this paper, only AI ethics for interaction with humans are considered, while AI ethics for interaction with non-human animals or with potential digital minds have been discussed elsewhere (Ziesche, 2021; Ziesche & Yampolskiy, 2019).

Structure

First, ten selected non-Western approaches to AI ethics are introduced. This is followed by a qualitative analysis of the proposals according to the research questions. The results are then presented as well as interpreted, which will be complemented by concluding recommendations.

Selected non-Western approaches to AI ethics

This review is guided by the circumstance that in the history of philosophy a variety of ethical schools have evolved, partly linked to religions, which constitute the foundation of current approaches to AI ethics.[2] Many of these ethical schools have developed tenets unknown to Western ethical approaches. For this article a selection of these concepts is introduced, which have been proposed to be included to universal AI ethics. These approaches can be categorized by being derived from certain belief systems or are covering specific regions or countries in the world and are partly overlapping.

Buddhism

One fundamental assumption in Buddhism that all sentient beings strive to **reduce pain**. Therefore, it has been requested that this goal has to be prominently incorporated in AI ethics. This is linked with the Buddhist concept of self-cultivation, which implies constant commitment, efforts and learning of all who are involved with AI to advance towards the goal of completely eliminating suffering (Hongladarom, 2020).

Another relevant concept of Buddhism is the **denial of a personal identity**. Instead, the Buddhist philosophy of mind uses the expression “anatta”, which means "non-self" and "holds that the notion of an unchanging permanent self is a fiction and has no reality" (Morris, 2006, p.51). In lieu thereof, a (sentient) being is defined by the following five so-called skandhas: Form, sensations, perceptions, mental activity or formations and consciousness. From this point of view, the issue of privacy protection, which plays a critical role in Western AI ethics, is “to chase a red-herring”, i.e., without personal identity privacy concerns are unfounded (Goodman, 2022).

Hinduism

Also, regarding Hinduism it has been noted that it is usually not represented in AI ethics, while it could add significant value (Sen, 2021). This concerns especially the Hindu tenet “**Dharma**”, which could be translated as “duty”, “action”, "religion" or “a sense of morality”.

According to the principle of dharma an action is righteous if the motive, the means adopted as well as the consequences of the action are righteous and in harmony. Unlike existing Western ethics, the dharma approach would address the issue that nowadays often only the motives of AI systems are non-maleficent, while the means and the consequences are frequently problematic, e.g., an AI system that is trained by the means of poor data that lead consequently to biases and discrimination (Sen, 2021).

[2] See here for overviews of some ethical schools: https://plato.stanford.edu/entries/african-ethics/, https://plato.stanford.edu/entries/ethics-chinese/ and https://plato.stanford.edu/entries/ethics-indian-buddhism/

Islam

Moreover, it has been proposed to integrate the Islamic legal doctrine "**Maqāṣid**" into an AI ethics framework as this may increase the chances for the acceptance of the global Muslim population (Raquib et al., 2022).

Maqāṣid is based on a hierarchy of three priorities: 1) Essentials are absolute necessities, 2) needs are less critical necessities, and 3) enhancements are optional, yet desirable. The essentials comprise the five objectives religion, life, progeny, property and intellect. It has been recommended that AI technology is based on a normative ethical framework, which seeks the values of Maqāṣid (Raquib et al., 2022).

Africa

A concept originating from Africa, which has been suggested to consider for AI ethics is "**ubuntu**". It "refers to a collection of values and practices that black people of Africa or of African origin view as making people authentic human beings. While the nuances of these values and practices vary across different ethnic groups, they all point to one thing – an authentic individual human being is part of a larger and more significant relational, communal, societal, environmental and spiritual world." (Mugumbate & Chereni, 2020, p. vi). In other words, this concept strengthens collectivism over individualism by highlighting the interdependence of humans and their responsibility for each other.

It has been stressed that countries from the Global South are hardly represented in the discourse about AI ethics. If Africa's ubuntu ethics were incorporated in AI ethics, this would strengthen values such as harmony, consensus, collective action as well as common good (Gwagwa et al., 2022). Moreover, it has been suggested that ubuntu could be harnessed to tackle the negative effects of automated decision-making systems and the economic, political and social arrangements that influence them by initially acknowledging humans as communal and social (Mhlambi, 2020).

China

Regarding AI in China a critical development has been the launch of the "New Generation Artificial Intelligence Development Plan", which details how China aims to become the world leader in AI by 2030. This plan also includes the goal that "by 2025 China will have seen the initial establishment of AI laws and regulations, ethical norms and policy systems, and the formation of AI security assessment and control capabilities."[3]

Linked to this, AI ethical guidelines have been endorsed by the Chinese National New Generation Artificial Intelligence Governance Professional Committee and comprise the following eight principles: Harmony and human-friendly, fairness and justice, inclusion and

[3] English translation: https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

sharing, respect for privacy, safe and controllability, shared responsibility, open collaboration as well as agile governance.[4]

It has been noted that these principles resemble Western approached to AI ethics. Yet, in reality the Chinese approach differs significantly from corresponding Western, e.g., EU, approaches, which can be explained through different philosophical traditions, cultural heritages, historical contexts and institutional disparities (Roberts et al., 2021, Fung & Etienne, 2022).

The EU approach is based on enlightenment values, such as individual freedom, equal rights and protection against abuses by the state. Instead, the Chinese approach is grounded in Confucian values, such as **virtuous government, harmonious society, social responsibilities and community relations** and with less focus on individualistic rights. Owing to the different foundations it was also stated that the EU principles concentrate on what AI must not do, thus AI risks, while the Chinese principles focus on opportunities of AI systems (Roberts et al., 2021, Fung & Etienne, 2022).

India

Fairness in machine learning has received in recent years due attention, yet focused on structural injustices prevailing in the West, such as gender and race and to some extent disability status, age and sexual orientation. However, there are various other **axes of discrimination** relevant to other geographies and cultures, which are hardly explored, yet nonetheless significantly contribute to biases in machine learning. In India existing algorithmic fairness assumptions are challenged, as they do not take into account further country-specific axes of discrimination such as caste, class and religion. This constitutes yet another reason that Western approaches to AI ethics are not applicable globally, but require contextualisation (Sambasivan et al., 2021).

Japan

While the Japanese philosophy has been also influenced by Buddhism and Confucianism (see above), another inherently Japanese concept is relevant for AI ethics, which is **techno-animism**. Techno-animism is an attitude to consider technology having human and spiritual characteristics, which is prevalent in Japan and can be traced to the Shinto religion (Jensen & Blok, 2013). Anecdotal evidence for this is that Japanese people to tend to have higher affinity with robots as they compare them with cherished manga characters, while in Western contexts robots are seen as soulless if not misanthropic as depicted in some movies. Therefore, a policy need has been expressed "for experiential sensitivity to objects, systems, synthetic personalities, emergent relationships, and the complex interactions that will surely emerge as emotion and affect-based systems grow more sophisticated in their capacities" (McStay, 2021, p. 19).

Another relevant Japanese concept is **ikigai**, which can be translated as "reason or purpose to live" (e.g., Kamiya, 1966). As scenarios, in which (a high number of) humans are devoid of

[4] English translation: http://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html

any ikigai are undesirable, they ought to be prevented and have been coined “i-risk scenarios”. It has been pointed out that developments in AI may lead to such i-risk scenarios, e.g., by taking over much more efficiently activities, which humans used to carry out day by day and considered them as their ikigai (Ziesche & Yampolskiy, 2020). While i-risk scenarios are both quite likely as well as very much unwanted, they are not yet reflected in any AI ethics.

Māori

It has been also attempted to design and evaluate AI from the perspective of the Māori, the indigenous people of New Zealand. Several Māori concepts, practices, and paradigms have been suggested as pertinent (Munn, 2023), out of which here the focus is on **mauri**. Mauri stands for the force or the quality of being alive. This concept has been characterized as “powerful in highlighting a holistic understanding of care for life” by bringing “together aspects of individual well-being, social support, good governance, and environmental sustainability (Munn, 2023). AI systems ought to preserve all kinds of mauri, which provides a useful umbrella notion for AI ethics, yet also indicates the complexity of AI ethics due to the interconnectedness of the concept of mauri.

Analysis

In several overviews of Western AI ethics lists of key ethical values and principles have been summarized. Examples are presented in Table 1 to provide a basis for analysing their compatibility with the presented non-Western concepts.

Source	**Jobin et al. (2019)**	**Hagendorff (2020)**	**High-Level Expert Group (2019)**	**Siau & Wang (2020)**
Title	*Eleven overarching ethical values*	*Top ten key issues*	*Seven key requirements that AI systems should meet in order to be trustworthy:*	*Five top factors of ethical frameworks*
Key issues	Transparency	Privacy protection	Societal and environmental well-being	Responsibility/Accountability
	Justice and fairness	Accountability	Diversity, non-discrimination and fairness	Privacy
	Non-maleficence	Fairness, non-discrimination, justice	Human agency and oversight	Transparency
	Responsibility	Transparency, openness	Privacy and data governance	Human Values/Do No Harm
	Privacy	Safety, cybersecurity	Technical Robustness and safety	Human Well-Being/Beneficence
	Beneficence	Common good, sustainability, well-being	Transparency	
	Freedom and autonomy	Human oversight, control, auditing	Accountability	

	Trust	Explainability, interpretabiliy		
	Dignity	Solidarity, inclusion, social cohesion		
	Sustainability	Science-policy link		
	Solidarity			

Table 1: Key issues in Western AI ethics

Table 1 also illustrates that these lists have overlaps. For example, transparency, privacy, responsibility and accountability as well as sustainability and well-being are mentioned in all four of them.

In the next step, the ten non-Western concepts introduced above are analysed according to the research question whether these aspects are compatible with Western AI ethics and could be merged, or whether these aspects are incompatible with Western AI ethics and hold potential for conflict. Table 2 provides an overview.

Concept	**Origin**	**Compatibility**	**Comment**
Pain reduction	Buddhism	Fully	This concept raises/strengthens awareness that AI systems may pose suffering risks (Althaus & Gloor, 2016) to sentient beings and is as such original and not yet represented in AI ethics.
Denial of a personal identity	Buddhism	Partially	This concept is related to a long-standing open philosophical question. An answer to this question does not seem to be necessary for AI ethics. AI systems should protect the privacy of those who wish it to be protected, while others may be indifferent in this regard.
Dharma	Hinduism	Fully	This concept raises/strengthens awareness that means and the consequences of AI systems are considered.
Maqāṣid	Islam	Fully	This concept raises/strengthens awareness that AI systems support that the essential necessities of all humans are satisfied.
Ubuntu	Africa	Partially	These collectivism approaches are not represented in Western AI ethics as such, yet also not necessarily incompatible with them, notwithstanding indeed similar to already covered issues such as solidarity, inclusion and social cohesion.
Virtuous government, harmonious society, social responsibilities and community relations	China	Partially	
Axes of discrimination	India	Fully	This concept raises/strengthens awareness that discrimination and bias have a variety of facets depending on cultures and context.

Techno-animism	Japan	Fully	This concept raises/strengthens awareness that acceptance and affinity are critical for humanity and AI to thrive together.
Ikigai	Japan	Fully	This concept raises/strengthens awareness of the importance of the purpose to live for people in an AI-dominated world.
Mauri	Māori	Fully	This concept raises/strengthens awareness that AI systems may pose an existential risk (Bostrom, 2002) to life on earth.

Table 2: Sample of non-Western concepts proposed to be used for AI ethics

In summary:

- Seven out of the ten concepts appear to be fully compatible with Western AI ethics, yet at the same time original and enriching.
- Three out of ten concepts appear to be partially compatible with Western AI ethics. These concepts deviate from Western approaches, nonetheless, they seem to be reconcilable Western AI ethics, while being original and enriching as well.
- Overall, these concepts fill gaps in existing Western AI ethics, partly critical gaps. For example, the three categories of risks existential, suffering and ikigai risks, also called x-, s- and i-risks, have not been adequately covered in AI ethics, yet could have severe consequences.

Conclusion

In conclusion, most of the introduced non-Western concepts are, at least to some extent, compatible with current AI ethics approaches. All these concepts are original and would provide beneficial enhancement as well as a broadening perspective to existing approaches. It is, therefore, recommended to incorporate them to create truly universal AI ethics. Apart from the inspiring content of these approaches, there is also a moral obligation to consider them, given that in Table 1 inclusion, diversity and non-discrimination are listed, yet non-Western concepts are hardly reflected so far.

Despite this recommendation the caveat has to be noted that there are still challenges anticipated: Somewhat related to AI ethics is the value alignment problem of AI, which is known for a while already as being very hard, even if only Western approaches are taken into account (e.g., Bostrom, 2014). This problem is, in brief, about ensuring that AI systems pursue goals and values, which are aligned with human goals and values, for which, firstly, all relevant values have to be precisely formulated and, secondly, these values have to be aggregated in a consistent manner. Both steps are complex, and will become more complex if non-Western approaches are included. Yet, this endeavour is nevertheless essential to leave no one behind, as has been summarized by Gabriel (2020, p.424-425) as follows: “It is to find a way of selecting appropriate principles that is compatible with the fact that we live in a diverse world, where people hold a variety of reasonable and contrasting beliefs about value. ... To avoid a situation in which some people simply impose their values on others, we

need to ask a different question: In the absence of moral agreement, is there a fair way to decide what principles AI should align with?“

It has to be stressed again that in this article only a selection of non-Western approaches to AI ethics has been presented. The motivation was to raise awareness for the issue that currently these approaches have been largely neglected, while this is not only morally wrong, but also means that at present AI ethics is missing out important and long-established ethical schools from large parts of the world. Therefore, it is recommended to widen this exercise. There are still a number of groups mostly from the Global South, from whom no proposals have been put forward how to represent their values in AI ethics guidelines, thus who are not represented at all currently.

Another remaining undertaking is the reverse analysis whether there are values in the existing Western AI ethics, which are incompatible with value systems of other parts of the world.

The paper concludes with a reiteration of the call to merge Western and non-Western approaches to AI ethics as far as they are compatible and to attempt to reconcile those aspects, which appear incompatible, as well as with a call to encourage groups who have not proposed any AI ethics yet to do so or offer them support in this regard.

References

Althaus, D., & Gloor, L. (2016). Reducing risks of astronomical suffering: a neglected priority. *Foundational Research Institute: Berlin, Germany*.

https://longtermrisk.org/reducing-risks-of-astronomical-suffering-a-neglected-priority/

Bostrom, N. (2002). Existential risks: Analyzing human extinction scenarios and related hazards. *Journal of Evolution and technology*, *9*.

https://nickbostrom.com/existential/risks

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*; Oxford University Press: Oxford, UK.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228*.

https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., ... & Terem, E. (2022). Worldwide AI Ethics: a review of 200 guidelines and recommendations for AI governance. *arXiv preprint arXiv:2206.11922*.

https://arxiv.org/ftp/arxiv/papers/2206/2206.11922.pdf

Fung, P., & Etienne, H. (2022). Confucius, cyberpunk and Mr. Science: comparing AI ethics principles between China and the EU. *AI and Ethics*, 1-7.

https://link.springer.com/article/10.1007/s43681-022-00180-6

Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds Mach, 30*, 411–437.

https://link.springer.com/article/10.1007/s11023-020-09539-2

Goodman, B. (2022). Privacy without persons: a Buddhist critique of surveillance capitalism. *AI and Ethics*, 1-12.

https://link.springer.com/article/10.1007/s43681-022-00204-1

Gwagwa, A., Kazim, E., & Hilliard, A. (2022). The role of the African value of Ubuntu in global AI inclusion discourse: A normative ethics perspective. *Patterns*, *3*(4), 100462.

https://www.sciencedirect.com/science/article/pii/S2666389922000423

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

https://arxiv.org/ftp/arxiv/papers/1903/1903.03425.pdf

High-Level Expert Group (2019). Ethics guidelines for trustworthy AI. European Commission.

https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html

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

Jensen, C. B., & Blok, A. (2013). Techno-animism in Japan: Shinto cosmograms, actor-network theory, and the enabling powers of non-human agencies. *Theory, Culture & Society*, *30*(2), 84-115.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.

https://arxiv.org/ftp/arxiv/papers/1906/1906.11668.pdf

Kamiya, M. (1966). Ikigai-ni-tsuite[On 755 ikigai]. Tokyo, Japan: MisuzuShyobou.

Kiemde, S. M. A., & Kora, A. D. (2022). Towards an ethics of AI in Africa: rule of education. *AI and Ethics*, 1-6.

McStay, A. (2021). Emotional AI, ethics, and Japanese spice: Contributing community, wholeness, sincerity, and heart. *Philosophy & Technology*, *34*(4), 1781-1802.

https://research.bangor.ac.uk/portal/files/39285253/2021_emotional_AI.pdf

Mhlambi, S. (2020). From rationality to relationality: ubuntu as an ethical and human rights framework for artificial intelligence governance. *Carr Center for Human Rights Policy Discussion Paper Series*, *9*.

https://carrcenter.hks.harvard.edu/files/cchr/files/ccdp_2020-009_sabelo_b.pdf

Morris, B. (2006). *Religion and anthropology: A critical introduction*. Cambridge University Press.

Mugumbate, J. R., & Chereni, A. (2020). Now, the theory of Ubuntu has its space in social work. *African Journal of Social Work*, *10*(1).

https://www.ajol.info/index.php/ajsw/article/view/195112

Munn, L. (2023). The five tests: designing and evaluating AI according to indigenous Māori principles. *AI & SOCIETY*, 1-9.

https://link.springer.com/article/10.1007/s00146-023-01636-x

Raquib, A., Channa, B., Zubair, T., & Qadir, J. (2022). Islamic virtue-based ethics for artificial intelligence. *Discover Artificial Intelligence*, *2*(1), 11.

https://link.springer.com/article/10.1007/s44163-022-00028-2

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI & society*, *36*, 59-77.

https://link.springer.com/article/10.1007/s00146-020-00992-2

Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021). Re-imagining algorithmic fairness in India and beyond. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 315-328).

https://arxiv.org/pdf/2101.09995.pdf

Sen, S. (2021). Can AI Ethics be informed by the Hindu concept of Dharma? Linkedin.

https://www.linkedin.com/pulse/can-ai-ethics-informed-hindu-concept-dharma-sujai-sen/

Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: ethics of AI and ethical AI. *Journal of Database Management (JDM)*, *31*(2), 74-87.

https://scholarsmine.mst.edu/cgi/viewcontent.cgi?article=1356&context=bio_inftec_facwork

UNESCO (2021d). Recommendation on the Ethics of Artificial Intelligence. Paris, UNESCO.

Ziesche, S. (2021). AI Ethics and Value Alignment for Nonhuman Animals. Special Issue “The Perils of Artificial Intelligence” of Philosophies, 6(2):31.

Ziesche, S. & Yampolskiy, R. V. (2019). Towards AI Welfare Science and Policies. Special Issue "Artificial Superintelligence: Coordination & Strategy" of Big Data and Cognitive Computing, 3(1):2.

Ziesche, S. & Yampolskiy, R. V. (2020). Introducing the Concept of Ikigai to the Ethics of AI and of Human Enhancements. In *Workshop on Ethics in AI & XR at 3rd International Conference on Artificial Intelligence & Virtual Reality*, 138-145.