Cultural Differences as Excuses? Human Rights and Cultural Values in Global Ethics and Governance of AI

Pak-Hang Wong¹

Received: 10 June 2020 / Accepted: 22 June 2020 / Published online: 23 July 2020

© The Author(s) 2020

Abstract

Cultural differences pose a serious challenge to the ethics and governance of artificial intelligence (AI) from a global perspective. Cultural differences may enable malignant actors to disregard the demand of important ethical values or even to justify the violation of them through deference to the local culture, either by affirming the local culture lacks specific ethical values, e.g., privacy, or by asserting the local culture upholds conflicting values, e.g., state intervention is good. One response to this challenge is the human rights approach to AI governance, which is intended to be a *universal* and *globally enforceable* framework. The proponents of the approach, however, have so far neglected the challenge from cultural differences or left out the implications of cultural diversity in their works. This is surprising because human rights theorists have long recognized the significance of cultural pluralism for human rights. Accordingly, the approach may not be straightforwardly applicable in “non-Western” contexts because of cultural differences, and it may also be critiqued as philosophically incomplete insofar as the approach does not account for the (non-) role of culture. This commentary examines the human rights approach to AI governance with an emphasis on cultural values and the role of culture. Particularly, I show that the consideration of cultural values is essential to the human rights approach for both philosophical and instrumental reasons.

Keywords Cultural diversity · Ethical pluralism · Human rights · Human dignity · Universal human interests · AI governance

✉ Pak-Hang Wong
wong@informatik.uni-hamburg.de

¹ Department of Informatics, Universität Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany

1 Introduction

In 2010, the Indian government stated that “India is not a particularly private nation. Personal information is often shared freely and without thinking twice. Public life is organized without much thought to safeguarding personal data” (quoted in Marda and Acharya 2014). The statement suggests that privacy is a foreign idea and being imported into India at most for instrumental purposes. In a similar vein, Bing Song (2018) has defended state intervention into private lives through China’s social credit system by noting that “there are different cultural expectations of the government in China than in other countries. China’s governance tradition of promoting good moral behavior goes back thousands of years.” The Chinese view is in marked contrast to the liberal view in which the value of individual autonomy significantly limits state intervention (see, e.g., Mill 1978).

The Supreme Court of India has subsequently ruled that privacy is a fundamental right guaranteed by the Indian constitution (Panday 2017). Also, the Chinese government has since explicitly called for the development of artificial intelligence (AI) technologies to safeguard societal security and respect human rights, which, in principle, protects Chinese citizens from unwarranted interference (Laskai and Webster 2019). These examples, however, are useful for illustrating the difficulties that *cultural differences* present to global ethics and governance of AI. In the worst-case scenarios, malignant actors could disregard commonly accepted ethical values or even justify violating them through deference to the *local culture*, either by affirming that the local culture lacks specific ethical values, e.g., privacy, or by asserting that the local culture upholds conflicting values, e.g., state intervention is good. Deference to cultural differences, therefore, can pose a serious challenge to the ethics and governance of AI from a global perspective, and it calls for an approach to AI ethics and governance that respects basic human values while also providing legal and ethical frameworks that are applicable and enforceable across different cultures.

The human rights approach to AI governance (see, e.g., Access Now 2018; IEEE 2018; Raso et al. 2018; Latonero 2018; AI HLEG 2019; ARTICLE 19 2019; Yeung et al. forthcoming) intends to be a *universal* and *globally enforceable* framework. The proponents of the human rights approach, however, have not fully engaged with the challenge from cultural differences or explored the implications of cultural diversity to the human rights approach in their works. This is surprising because human rights theorists have long recognized the significance of cultural pluralism for human rights. For instance, the idea of human rights has been faulted for being inherently “Western,” and so it may not be straightforwardly applicable to “non-Western” cultures (see, e.g., Panikkar 1982; Bell 1996; Metz 2012). There is also a longstanding philosophical debate over the role of culture in the normative justification of human rights (see, e.g., Cohen 2004).

The aim of this commentary, therefore, is to foreground the importance of cultural values and the role of culture in the human rights approach to AI governance. I start with a brief overview of the human rights approach, and then outline some problems the approach faces. I argue that the conversation about human rights and AI technologies has so far overlooked the question of the nature of “human rights,” which contributes to the omission of cultural values. I then discuss two understandings of human rights, i.e., moral conceptions and political conceptions of human rights, and

illustrate how they require the human rights approach to take cultural values seriously. In the concluding remarks, I draw some lessons from my discussion for engaging with cultural values for the human rights approach and for global ethics and governance of AI technologies more generally.

2 A Sketch of the Human Rights Approach to AI Ethics and Governance

There is an increasing emphasis on human rights in the conversation about the ethical and governance issues raised by AI technologies. Scholars and practitioners have made explicit reference to the major European and international human rights documents in their works on AI governance, including, for example, the Universal Declaration of Human Rights (UDHR), the International Covenant on Civil and Political Rights (ICCPR), the International Covenant on Economic, Social and Cultural Rights (ICESCR) (e.g., Access Now 2018; Raso et al. 2018; Latonero 2018), the UN Guiding Principles on Business and Human Rights (GPs) (e.g., Raso et al. 2018; ARTICLE 19 2019), and the European Convention on Human Rights (ECHR) (e.g., Yeung et al. forthcoming). I shall categorize these works under the label of the human rights approach to AI ethics and governance, or HRA for short. Central to HRA is the *explicit* reference to major European and international human rights instruments as the *normative* standard for the development and implementation of AI technologies.

Proponents of HRA draw attention to the wide acceptance of the UDHR and related documents and argue that human rights provide a global normative standard, as well as a shared normative language, to assess and deliberate on the potential social, ethical, and political implications of AI technologies. They also maintain that international human rights instruments supply enforceable mechanisms of accountability and redress required by genuine governance of AI technologies. Karen Yeung et al. (forthcoming) point out that there are well-established institutions and organizations at the international and national level to promote human rights and to monitor, prevent, and respond to human rights violations, which HRA can utilize in practice. In short, the appeal to human rights in HRA serves two purposes, i.e., (i) it offers a normative standard to identify, anticipate, and evaluate the harmful (or beneficial) impacts of AI technologies and (ii) a set of legal and institutional measures to prevent, mitigate, and rectify the harm caused by them based on the existing legal and institutional frameworks. For example, Mark Latonero (2018, 13–14) refers to UDHR Article 12 (and ICCPR Article 17): “No one shall be subject to arbitrary interference with his privacy, family, home or correspondence [...]” to ground the fundamental normative value of privacy and suggests that treating privacy as a fundamental human right can strengthen privacy considerations in existing industry norms and technical standards. Likewise, Filippo Raso et al. (2018) evaluate various applications of algorithms in major areas of life based on how they affect the list of human rights specified in UDHR.

More importantly, HRA is considered to be readily applicable to “non-Western” contexts. Consider the case of China’s social credit system. The system aims to collect personal data from Chinese citizens, and classify their behaviors into good and bad behaviors, and then reward the individuals when they behave in accordance with what the Chinese government considers to be good and punishes them for behaviors that are

deemed bad via automated systems.¹ HRA will critique the system as legally and morally impermissible for an excessive intrusion of private lives, which constitutes a violation of UDHR Article 12 and other Articles concerning the rights to freedom, e.g., Article 18, 19, and 20. In this way, the UHDR and related documents serve as a *universal* normative standard in HRA, which is supposedly applicable across nations and cultures.

3 Questioning the Human Rights Approach

It is clear from the above summary that HRA does offer an explicit and established normative standard to examine the normative issues of AI technologies. In the current conversation about human rights and AI technologies, however, the proponents of HRA have not sufficiently acknowledged the contested nature of human rights and the extent of disagreement over their interpretation. Indeed, most of the works on HRA mentioned earlier simply assume the universality and (global) legitimacy of human rights, even though this has been challenged in different ways.²

There are worries about the global power asymmetry in shaping the human rights agenda and practices, thereby turning human rights into a tool to serve the interests of power (see, e.g., Gilabert 2018). There are also issues with conflicting interpretations of specific human rights and their implementation and the criteria to evaluate and settle conflicts between the different interpretations (see, e.g., Macioce 2016, 2018).³ These problems, of course, are not decisive against the idea of human rights, but they do hint at the need to include cultural values in the discussion of human rights. Pablo Gilabert (2018) argues that among other criteria, attentiveness to local differences, human interpretative fallibility, and humility and multiplicity in the interpretation of human rights are essential in countering undue power and the misuse of human rights. Also, reference to local values can help to resolve disputes over conflicting interpretations of human rights in practice (Macioce 2016, 2018). In addition, researchers in intercultural information ethics have demonstrated the interplay among information technology, cultures, and ethical values (see, e.g., Capurro 2005), and they have also argued that there is a plurality of *reasonable* moralities from different cultures for assessing the ethical issues of information technology and warned against ethical universalism for its propensity to dogmatism and intolerance (see, e.g., Ess 2006; Capurro 2007). As such, proponents of HRA should be wary of the missing discussion on the contested nature of human rights and the possible disagreement over their interpretation and, relatedly, the implications of a plurality of reasonable moralities.

¹ This is, by no means, a sufficient analysis of China's social credit system. Instead, this brief analysis only intends to illustrate how HRA can be applied in a cross-cultural setting. For a nuanced analysis of China's social credit system, see Engelmann et al. (2019) and Liu (2019).

² For exceptions, see Latonero (2018) and IEEE (2018).

³ It is worth noting that Yeung et al. (forthcoming) refer explicitly to the ideal of democracy to defend HRA. However, the reference to democracy may be problematic insofar as HRA aims to be universal and globally enforceable, because then either HRA is inapplicable to non-democratic states, or it regard democracy as the *only* form of political system that is legitimate and rejects the legitimacy of non-democratic states, see Wong (2016).

Proponents of HRA have also largely overlooked the question of the nature of human rights in their works. While some accounts of HRA have made explicit their normative ground or their understanding of human rights, e.g., Latonero (2018) mentioned human dignity, Yeung et al. (forthcoming) referred to the value of democracy, and Raso and colleagues adopted a “legal conception of human rights” (2018, 12), they have not further explored the normative and pragmatic implications of using these normative grounds and understandings of human rights. Yet, this question is pertinent to the present discussion as the different understandings of human rights assume different stances toward the relevance of cultural values.

4 Cultural Values and the Two Conceptions of Human Rights in the Human Rights Approach

A clarification of the understandings of human rights in HRA enables a better appreciation of the questions raised by cultural differences by foregrounding the relation between the normative basis of human rights and cultural values. The two major accounts of human rights I shall examine are the moral conceptions and the political conceptions of human rights.

As Rowan Cruft, S. Matthew Liao, and Massimo Renzo succinctly summarize, the key disagreement between them is that according to the moral conceptions of human rights, “[human rights] are (a) moral rights that (b) all human beings possess (c) at all times and in all places (d) simply in virtue of being human and (e) the corresponding dutybearers are all able people in appropriate circumstances” (Cruft et al. 2015, 4), whereas the “political conceptions [of human rights] argue that human rights are not based on certain features of humanity; rather, the distinctive nature of human rights is to be understood in light of their role or function in modern international political practice” (Cruft et al. 2015, 6). Instead of summarizing various formulations of moral and political conceptions of human rights, I shall refer to the accounts of human rights by John Tasioulas and John Rawls, respectively, to illustrate the differences the two conceptions make to HRA.

4.1 Moral Conceptions of Human Rights and the Human Rights Approach

John Tasioulas (see, e.g., 2012, 2013, 2015) offers one of the most comprehensive moral conceptions of human rights. In his pluralist approach to the grounds of human rights, Tasioulas holds that “human dignity and universal human interests are equally fundamental grounds of human rights, characteristically bound together in their operation” (Tasioulas 2015, 53–54). Tasioulas defends “the human nature conception of human dignity”, in which the normative value of human dignity derives from a variety of capacities and features that constitute the species of *human beings*, and argues that this view of human dignity guarantees an equality of basic moral status and a fundamental level of respect among individuals *qua* human beings (2015, 54–55). For human interests, they are the things required for or which contribute to *human* well-being, and, by universal human interests, Tasioulas refers to human interests that are “*objective, standardized, pluralistic, open-ended, and holistic* in character” (2015, 50).

In the conversation of human rights and AI technologies, the idea of human dignity finds its place in Latonero's account, where he states that "[i]n bridging AI and human rights, what's at stake is human dignity" (2018, 5; also, see AI HLEG 2019; Yeung et al. [forthcoming](#)). Latonero's and other accounts that invoke human dignity or universal human interests as normative justifications can be viewed as assuming the moral conception.

There are various critiques of the moral conception, in particular of the idea of human dignity for being vague and for lacking cross-cultural legitimacy. For present purposes, I focus on the criticism about the cross-cultural legitimacy of human dignity.

Peimin Ni (2014) argues that early Confucians regard *the potentiality to become morally excellent* as the defining feature of human beings and that this potentiality can serve as the foundation of human dignity in Confucianism. However, he also points out that early Confucians believe individuals can lose this potentiality through a sustained failure to fulfill their moral responsibility. For Ni, "human dignity [is] an achievement, a demand one develops from the world of concrete feelings within, a self-realization obtained through one's efforts in "establishing" and "promoting" others" (2014, 195). This account differs markedly from the common account of human dignity in that it insists that being a human being does not in itself guarantee individuals an equality of basic moral status and a fundamental level of respect.

Ni's account challenges the common account as the normative ground for *universal* human rights and therefore also challenges Latonero's and similar accounts of HRA that are based on the common account. If HRA is based on the common account of human dignity, and when the common account is not shared by other cultural traditions, e.g., Ni's Confucian account of human dignity, then HRA lacks normative justification in those cross-cultural settings, or HRA has to reject alternative understandings of human dignity in other reasonable moralities. Both options are problematic. The first option rejects HRA as a genuinely universal and globally enforceable framework. For the second option, as Simon Caney notes, an outright dismissal of other moralities without first engaging with them can only be "a form of philosophical arrogance" and would leave us "epistemically handicapped" (2000, 60).

Similar arguments can be made about universal human interests. Recall the Indian government's claim that privacy is *not* a value in the Indian value system, or, recall Song's claim that citizens in China have little expectation (or, interest) in non-interference of the state. In both cases, the aim is to deny some human interests as *universal* and assert that those interests are not a sound normative justification. Possible cultural variation in human interests therefore calls into question universal human interests as the normative ground for human rights and for HRA.

It should be clarified that the discussion above is not a critique against HRA in toto, but only against the accounts of HRA that depend on the moral conception *without* taking cultural differences seriously. The discussion demonstrates the need to engage with other reasonable moralities in constructing the normative foundation of the approach. Caney (2000) has rightly noted that successful inclusion of the values in other reasonable moralities can strengthen the legitimacy of human rights by ensuring the people who espouse reasonable moralities in different cultures assent to them through their *own* values. It is also helpful to remember that major cultural traditions in the world have *within* themselves rich normative resources, and the cross-cultural applicability of HRA can be much strengthened by drawing upon them. For instance,

pace Ni, proponents of HRA can draw from Irene Bloom's (1998) Confucian account of human dignity, which stresses instead the potentiality to become morally excellent as shared by *all* individuals, to defend the normative significance of human dignity for human rights and for HRA.

Or, returning to the two cases mentioned at the beginning. The Indian government's claim that privacy is *not* an interest of people has been forcefully countered by efforts to demonstrate privacy as a fundamental value in classical Hindu law (Ashesh and Acharya 2014) and Islamic law (Marda and Acharya 2014). Recent works in Confucian political philosophy have also demonstrated the need for a virtuous (Confucian) government to restrict itself in interfering with the lives of people in order to allow people to genuinely acquire Confucian virtues through their *own* acts (see, e.g., Angle 2015). Accordingly, by drawing on a plurality of reasonable moralities, HRA can normatively evaluate the violation of privacy by or with AI technologies in the Indian context through the Hindu and/or Islamic understanding of the value of privacy, and they can also remain critical of China's social credit system through the view of virtuous Confucian government.

4.2 Political Conceptions of Human Rights and the Human Rights Approach

There are justified concerns about the moral conceptions of human rights given the plurality of reasonable moralities and the incommensurable values those moralities command.⁴ The political conceptions of human rights are formulated in part as a response to these concerns. Instead of grounding human rights in a metaphysical or theological view of human nature or human well-being, the political conceptions ground human rights by their functions. For instance, John Rawls has developed a political conception of human rights, in which human rights serve as a standard for sanction and intervention and for membership in good standing in the international community (1999, 79–80). More generally, Allen Buchanan states that the function of human rights “is to provide a set of universal standards, in the form of international law, whose primary purpose is to regulate the behavior of states toward individuals under their jurisdiction, considered as social individuals, and for their own sakes” (2013, 86). This seems to be the idea of human rights assumed by Filippo Raso and colleagues when they state that “[w]e view human rights in terms of the binding legal commitments the international community has articulated” (2018, 8).

The political conceptions overcome the justified concerns about the plurality and incommensurability of reasonable moralities by grounding human rights in their functions that are independent of specific cultures and would be acceptable to those who share the need for a well-functioning international order. However, the *content* of human rights that could be derived from their functions is foreseeably minimal, given the disagreement between sovereigns in the international community, e.g., Rawls' own list of human rights includes only “freedom from slavery and serfdom, liberty (but not equal liberty) of conscience, and security of ethnic groups from mass murder and genocide” (1999, 79). It may, therefore, seem doubtful whether the political conception can support the extended list of human rights that HRA intends to utilize in the normative evaluation of AI technologies.

⁴ It is beyond the scope of this commentary to defend the plurality of reasonable moralities and incommensurability between different ethical traditions. For a comprehensive defense of the view, see Wong (2009).

Here, the proponents of HRA, who favor a political conception of human rights, can expand the content of human rights by engaging with cultural values. As Joshua Cohen (2004) argues, the resulting content of human rights can be more than minimal if we identify human rights at the juncture of convergence among cultures. He distinguishes between *substantive minimalism* that requires human rights to be grounded in a set of de facto overlapping values in various cultural traditions from *justificatory minimalism* that maintains human rights should “be independent of particular philosophical or religious theories that might be used to explain and justify its content” (2004, 193) such that they can be justified and accepted by various reasonable moralities. Cohen’s justificatory minimalism entails that the proponents of HRA need to look into the values in other reasonable moralities if they are to justify a broad(er) range of human rights for normative evaluation of AI technologies in a cross-cultural context.

In fact, Cohen’s justificatory minimalism places the burden of proof on those who want to invoke cultural differences to disregard or violate the existing set of human rights. Insofar as accepting the existing set of human rights is a condition for membership in good standing in the international community, it is the deviation from them that should be justified. Here, justificatory minimalism does not entail that the existing set of human rights *always* trumps cultural values. Instead, it entails that the existing set of human rights is the *normative default* but nonetheless be open for reflection and deliberation from various cultural perspectives. In other words, the accounts of HRA based on a political conception also need to be open and responsive to justifications for divergence from the existing set of human rights in other cultures.

Alternatively, Benedict S. B. Chan (2019) has proposed to justify human rights by looking at the positive consequences they produce in practice *qua* international human rights. He illustrates with the example of the right to privacy. According to Chan, privacy is, and should be, included as a human right *because of* its importance to the society and globally and also *because of* the good consequences provide to individuals with and through it. So construed, insofar as the proponents of HRA can demonstrate the positive consequences of the rights in question, they can use them as the normative standard to evaluate AI technologies. Chan’s proposal is particularly useful to supplement HRA in that consequential evaluation enables the proponents of HRA to decide and defend *what human rights can and should be included* on the basis of their positive consequences, which, in turn, is important in countering an unnecessary proliferation of human rights for AI technologies. As Chan acknowledges, what constitutes good consequence in the consequential evaluation remains open to debate. I shall add to his view that cultural values ought to be included in the evaluation. I contend that the *actualization* of the good consequences of human rights, in whatever way one formulates them, depends significantly on local support. Accordingly, engaging with cultural values is important, albeit instrumentally, to HRA supported by the consequential evaluation.

5 Concluding Remarks: Cultural Values and Global Ethics and Governance of AI

The discussion so far has elaborated two understandings of human rights for HRA and shown that engaging with cultural values is essential to HRA either for *normative reasons* (for HRA based on the moral conception of human rights) or for *instrumental*

reasons (for HRA based on the political conception of human rights). To conclude, I draw some lessons from the prior discussion for engaging with cultural values for HRA and for global ethics and governance of AI technologies more generally.

Firstly, the discussion demonstrates that normative standards for the *global* ethics and governance of AI technologies should be viewed not as a pre-determined endpoint but as an on-going process of negotiation and construction, and thus ought to be *open* and *responsive* to cultural values. Secondly, the requirement of openness and responsiveness demands scholars and practitioners of AI ethics and governance to think *together with* the cultural others when deciding whether specific normative standards are appropriate for the evaluation of AI technologies and why they are appropriate in a cross-cultural setting. Doing so can mitigate the misuse of cultural differences, and it can also enrich the normative foundation of HRA and other approaches for the *global* ethics and governance of AI. For example, the normative justification of the right to privacy from the Hindu and Islamic traditions can be *added* to HRA and other approaches to global ethics and governance of AI technologies, and the same is true of the Confucian justification for state non-interference. Finally, *local* support seems to be indispensable to give teeth to the normative evaluation of AI technologies. Local support can best be acquired by deferring to the local community's values to justify the rights and other normative values or by proving to them that upholding the rights and normative values will result in good consequences *for them*. This can only be done if the proponents of HRA and other approaches to global ethics and governance of AI technologies are willing to work *with* and *in* other cultures.

Funding Information Open Access funding provided by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Access Now. (2018). Human rights in the age of artificial intelligence. Available Online at: <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf>.
- AI HLEG [High-Level Expert Group on Artificial Intelligence]. (2019). Ethics guidelines for trustworthy AI. Available Online at: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.
- Angle, S. (2015). Virtue ethics, the rule of law, and the need for self-restriction. In B. Bruya (Ed.), *The philosophical challenge from China* (pp. 159–182). Cambridge: MIT Press.
- ARTICLE 19. (2019). Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence. Available Online at: https://www.article19.org/wp-content/uploads/2019/04/Governance-with-teeth_A19_April_2019.pdf.
- Ashesh, A. & Acharya, B. (2014). Locating constructs of privacy within classical Hindu law. The Centre for Internet and Society. <https://cis-india.org/internet-governance/blog/loading-constructs-of-privacy-within-classical-hindu-law>

- Bell, D. A. (1996). The East Asian challenge to human rights: reflections on an East West dialogue. *Human Rights Quarterly*, 18(3), 641–667.
- Bloom, I. (1998). Fundamental intuition and consensus statements: Mencian Confucianism and human rights. In W. T. de Bary (Ed.), *Confucianism and human rights* (pp. 94–116). New York: Columbia University Press.
- Buchanan, A. (2013). *The heart of human rights*. New York: Oxford University Press.
- Caney, S. (2000). Human rights, compatibility and diverse cultures. *Critical Review of International Social and Political Philosophy*, 3(1), 51–76.
- Capurro, R. (2005). Privacy. An intercultural perspective. *Ethics and Information Technology*, 7(1), 37–47.
- Capurro, R. (2007). Intercultural information ethics. In R. Capurro, J. Frühbauer, & T. Hausmanninger (Eds.), *Localizing the Internet. Ethical aspects in intercultural perspective* (pp. 21–38). Fink: Munich.
- Chan, B. S. B. (2019). A preliminary consequential evaluation of the roles of cultures in human rights debates. *Filosofia Theoretica: Journal of African Philosophy, Culture and Religions*, 8(1), 163–182.
- Cohen, J. (2004). Minimalism about human rights: the most we can hope for? *Political Philosophy*, 12(2), 190–213.
- Cruft, R., Liao, S. M., & Renzo, M. (2015). The philosophical foundations of human rights: An overview. In R. Cruft, S. M. Liao, & M. Renzo (Eds.), *The philosophical foundations of human rights* (pp. 1–41). Oxford: Oxford University Press.
- Engelmann, S., Chen, M., Fischer, F., Kao, C., & Grossklags, J. (2019). Clear sanctions, vague rewards: how China's social credit system currently defines “good” and “bad” behavior. In *Proceedings of the Conference on Fairness, Accountability, and Transparency FAT* '19* (pp. 69–78). New York: Association for Computing Machinery.
- Ess, C. (2006). Ethical pluralism and global information ethics. *Ethics and Information Technology*, 8(4), 215–226.
- Gilbert, P. (2018). Reflections on human rights and power. In A. Etinson (Ed.), *Human rights: moral or political?* (pp. 375–399). Oxford: Oxford University Press.
- IEEE [The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems]. (2018). Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems. IEEE. <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf>.
- Laskai, L., & Webster, G. (2019). Translation: Chinese expert group offers ‘governance principles’ for ‘responsible AI’. New America, June 17, 2019. Available Online at: <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-expert-group-offers-governance-principles-responsible-ai/>.
- Latonero, M. (2018). Governing artificial intelligence: upholding human rights and dignity. Data & Society. Available Online at: https://datasociety.net/pubs/governing_ai.pdf.
- Liu, C. (2019). Multiple social credit systems in China. *Economic Sociology_The European Electronic Newsletter*, 21(1), 22–32.
- Macioce, F. (2016). Balancing cultural pluralism and universal bioethical standards: A multiple strategy. *Medicine, Health Care and Philosophy*, 19(3), 393–402.
- Macioce, F. (2018). Charters of rights and bioethical principles: a multicultural challenge. In C. Caporale & I. R. Pavone (Eds.), *International biolaw and shared ethical principles: the universal declaration on bioethics and human rights* (pp. 40–57). London: Routledge.
- Marda, V., & Acharya, B. (2014). Identifying aspects of privacy in Islamic law. The Centre for Internet and Society. <https://cis-india.org/internet-governance/blog/identifying-aspects-of-privacy-in-islamic-law>.
- Metz, T. (2012). African conceptions of human dignity: Vitality and community as the ground of human rights. *Human Rights Review*, 13(1), 19–37.
- Mill, J. S. (1978). *On liberty*. Indianapolis: Hackett Publishing.
- Ni, P. (2014). Seek and you will find it; let go and you will lose it: exploring a Confucian approach to human dignity. *Dao*, 13(2), 173–198.
- Panday, J. (2017). India's Supreme Court upholds right to privacy as a fundamental right—and it's about time. Electronic Frontier Foundation, August 28, 2017. Available Online at: <https://www EFF.org/deepinks/2017/08/indias-supreme-court-upholds-right-privacy-fundamental-right-and-its-about-time>.
- Panikkar, R. (1982). Is the notion of human rights a Western concept? *Diogenes*, 30, 75–102.
- Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. (2018). *Artificial intelligence & human rights: opportunities & risks*. Berkman Klein Center Research Publication No. 2018-6. Available Online at <https://ssrn.com/abstract=3259344>.
- Rawls, J. (1999). *The law of peoples: with “the idea of public reason revisited”*. Cambridge: Harvard University Press.

- Song, B. (2018). The West may be wrong about China's social credit system. *The Washington Post*, November 29, 2018. Available Online at: <https://www.washingtonpost.com/news/world/wp/2018/11/29/social-credit/>.
- Tasioulas, J. (2013). Human dignity and the foundations of human rights. In C. McCrudden (Ed.), *Understanding human dignity* (pp. 291–312). Oxford: Oxford University Press.
- Tasioulas, J. (2012). On the nature of human rights. In G. Ernst & J. Heilinger (Eds.), *The philosophy of human rights: contemporary controversies* (pp. 17–59). Berlin: de Gruyter.
- Tasioulas, J. (2015). On the foundations of human rights. In R. Cruft, S. M. Liao, & M. Renzo (Eds.), *The philosophical foundations of human rights* (pp. 45–70). Oxford: Oxford University Press.
- Wong, D.-B. (2009). *Natural moralities: a defense of pluralistic relativism*. New York: Oxford University Press.
- Yeung, K., Howes, A., & Pogrebna, G. (forthcoming). AI governance by human rights-centred design, deliberation and oversight: An end to ethics washing. In M. Dubber & F. Pasquale (Eds.), *The Oxford handbook of AI ethics*. New York: Oxford University Press Available Online at: <https://ssrn.com/abstract=3435011>.

Anonymized Reference:

- Wong, P.-H. (2016). Responsible innovation for decent nonliberal peoples: a dilemma? *Journal of Responsible Innovation*, 3(2), 154–168.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.