Just Hierarchy and the Ethics of Artificial Intelligence: Two Approaches to a Relational Ethic for Artificial Intelligence

Qin Zhu
Virginia Tech, US

ABSTRACT. Dominant approaches to the ethics of artificial intelligence (AI) systems have been mainly based on individualistic, rule-based ethical frameworks central to Western cultures. These approaches have encountered both philosophical and computational limitations. They often struggle to accommodate remarkably diverse, unstable, complex contexts of human-AI interactions. Recently there has been an increasing interest among philosophers and computer scientists in building a *relational* approach to the ethics of AI. This article engages with Daniel A. Bell and Pei Wang's most recent book *Just Hierarchy* and explores how their theory of just hierarchy can be employed to develop a more systematic account for relational AI ethics. Bell and Wang's theory of just hierarchy acknowledges that there are morally justified situations in which social relations are not equal. Just hierarchy can exist both between humans and between humans and machines such as AI systems. Therefore, a relational ethic for AI based on just hierarchy can include two theses: (i) AI systems should be considered *merely* as tools and their relations with humans are hierarchical (e.g. designing AI systems with lower moral standing than humans); and (ii) the moral assessment of AI systems should focus on whether they help us realize our role-based moral obligations prescribed by our social relations with others (these relations often involve diverse forms of morally justified hierarchies in communities). Finally, this article will discuss the practical implications of such a relational ethic framework for designing socially integrated and ethically responsive AI systems.

KEYWORDS. Artificial intelligence, just hierarchy, non-Western ethical perspectives, relational ethics

I. INTRODUCTION

Dominant approaches to the ethics of artificial intelligence (AI) systems have been mainly based on individualistic, rule-based ethical frameworks central to Western cultures. In this sense, the moral quality of an action taken by an artificial entity depends solely on its consistency with ethical principles (e.g. categorical imperatives, the principle of utility).

ETHICAL PERSPECTIVES 30, no. 1 (2023): 59-000.

doi: 10.2143/EP.30.1.0000000

These approaches have encountered both philosophical and computational limitations. They often struggle to accommodate remarkably diverse, unstable, complex contexts of human-AI interactions (Vallor 2016). Most significantly, to accommodate the extensively diverse contexts in which robots are situated, rule-based moral algorithms need to integrate a vast number of conditionals or 'exceptional cases' to which predetermined moral rules may not be applicable (Klincewicz 2017). These conditionals need to be frequently searched and revised while robots are completing tasks in specific contexts. Doing so can be extremely time and energy consuming. To address these limitations with rule-based approaches, philosophers have been exploring alternative approaches to AI ethics. For instance, a few scholars have discussed possibilities to construct a virtue-based framework for AI ethics (Hagendorff 2022; Vallor 2016). Nevertheless, such a virtue-based framework provides little or no guidance on how to appropriately design AI systems. Furthermore, there is a potential worry that Western virtues will assume a more dominant role in such a framework given that non-Western cultures have much less been represented in global AI ethics discourse. More recently there has been an increasing interest among philosophers and computer scientists in building a *relational* approach to the ethics of AI (Coeckelbergh 2010). Such an approach argues that the moral status, responsibilities, and competences of AI entities (e.g. robots) are determined by how these entities "[…] appear to us in the context of the concrete human-robot relation and the wider social structures in which the relation is embedded" rather than "the *real* ontological features of these entities" (Coeckelbergh 2010, 214). Relationships between AI entities and humans rather than the actual reasoning architecture of AI entities matter more in evaluating the moral implications of these AI systems.

The present contribution engages with Daniel A. Bell and Pei Wang's most recent book *Just Hierarchy* (Bell and Wang 2020) and explores how their theory of just hierarchy can be employed to develop a more systematic account for *relational* AI ethics. Bell and Wang's book briefly discusses some preliminary implications of the just hierarchy theory for understanding AI

ethics (part of Chapter Five). This article attempts to further expand the discussion of the relevance of the just hierarchy for AI ethics research and construct a more systematic, theoretical framework for AI ethics based on the just hierarchy theory by incorporating empirical findings from AI research. To do so, I adopt a critical engagement approach. In other words, while conceptualizing a relational ethic for AI based on Bell and Wang's work, I also briefly mention potential concerns raised by critics and discusses how Bell and Wang would potentially react to these criticisms. I argue that such a critical engagement approach is helpful for not only developing a more robust theory but also making such a theory more plausible when it is employed to inform practice such as ethical AI design.

In general, Bell and Wang's theory of just hierarchy acknowledges that there are morally justified situations in which social relations are not equal. Just hierarchy can exist both between humans and between humans and machines such as AI systems. More specifically, by leveraging Bell and Wang's work on just hierarchy, I conceptualize that there are two approaches through which AI technologies can contribute to morally justified and yet necessary social hierarchies: (i) AI systems should be considered *merely* as tools and their relations with humans are hierarchical (e.g. designing AI systems with lower moral standing than humans); and (ii) the moral assessment of AI systems should focus on whether they help us realize our role-based moral obligations prescribed by our social relations with others (these relations often involve diverse forms of morally justified hierarchies in communities). Finally, I briefly discuss the practical implications of such a relational ethic framework for designing socially integrated and ethically responsive AI systems in broader social and cultural contexts.

II. Just Hierarchy and Hierarchical Relationships

The term hierarchy has been widely criticized in not only the humanities and social sciences but also more recently in the fields of engineering and science, especially fields that involve emerging technologies such as AI,

robotics, and big data (Benjamin 2019). According to Western ethics and political philosophy, ranking individuals and groups based on particular cultural attributes (e.g. gender, race, ethnicity) is *always* morally unjustified and thus problematic. For instance, it is morally unquestionable to limit minority students from accessing science and engineering programmes as they often 'demonstrate' lower grades or GPAs and thus they are less qualified for these programmes. Nevertheless, we are also aware that not all social relations in society are (or should be) *always* equal. We sometimes value more experienced colleagues than others in the workplace. We admire those with well-developed moral virtues more than petty persons. Arguably we sometimes do not assign equal moral considerations to strangers and those with whom we share close relationships (e.g. parents, children, friends).

Bell and Wang (2020) invite us to critically examine the limitations of dominant Western, liberal democratic criticisms of social hierarchy. Instead of debating whether a good society should or should not have social hierarchies, they suggest that the real choice we need to make is rather "[…] between a society with unjust hierarchies that perpetuate unjust power structures and one with just hierarchies that serve morally desirable purposes" (2020, 13-14). According to them, a major reason why we reject most traditional hierarchies is not simply that hierarchies *per se* are wrong but that most traditional hierarchies are not morally justifiable and in fact violate fundamental human rights. There are certain forms of social hierarchies such as those based on noble birth, race, sex, and religious beliefs that are morally unjustifiable. In contrast, there are other forms of social hierarchies among family members, citizens, states, humans and animals, and humans and machines are morally defensible (Bell and Wang 2020).

In Confucian ethics, there are five cardinal role-based, often hierarchical relationships are of critical importance and these relationships are those between parents and children, husband and wife, older and younger siblings, rulers and ministers, and friends (Cottine 2020). The five relationships

belong to three social spheres: (i) family sphere: the parent-child relationship, the husband-wife relationship, and the relationship between siblings; (ii) intermediary sphere: friendship; and (iii) the social/political sphere: the ruler and minister relationship. Even among the five Confucian cardinal relationships, there is a hierarchy. Bell and Wang (2020) argue that Confucian ethics defends the value of partiality. According to them,

> Our ethical obligations are strongest to those with whom we have personal relationships, and they diminish in intensity as the father we go from those relationships. We do have an obligation to extend love beyond intimates, but it is not expected that the same degree of emotions and responsibilities will extend to strangers, and even less so to nonhuman forms of life (2020, 190).

Cottine (2020) argues that family relations are foundational for individual moral development and state governance. Arguably, being a filial child provides a paradigmatic case for being a loyal minister. Most of these five cardinal relationships are intimate relationships and the ruler-minister relationship is analogous to the parent-child relationship.

In their book, Bell and Wang (2020) further expand the traditional five Confucian cardinal relationships to include hierarchical relationships more than simply intimate relationships. They discuss five different kinds of social hierarchies that can be morally justifiable, namely those between: (i) intimates, (ii) citizens, (iii) states, (iv) humans and animals, and (v) humans and machines. Given that human-technology interaction (HCI) mainly addresses issues related to how humans interact with technologies and such interactions affect human-human relationships, the present contribution mainly addresses the social hierarchies articulated by Bell and Wang (2020) in two different senses: (i) social hierarchies between humans and machines; and (ii) social hierarchies between humans mainly including our relationships with other humans and sometimes animals. The social hierarchies between states will not be a major focus of this article.

Therefore, a relational ethic for AI informed by the just hierarchy theory can include at least two theses: (i) AI systems should be considered merely as tools and their relations with humans are hierarchical (e.g. designing AI systems with lower moral standing than humans); and (ii) the moral assessment of AI systems should focus on whether they help us realize our role-based moral obligations prescribed by our social relations with others (these relations often involve diverse forms of morally justified hierarchies in communities).

III. Two Approaches to a Relational Ethic for AI

AI Systems as Tools

The first approach to a relational ethic for AI based on the just hierarchy theory is to consider AI systems *merely* as tools to serve human needs. Bell and Wang's reasoning in this regard has been strongly influenced by the Marxist philosophy of technology, including Chinese experience of implementing such a philosophy in the governance of technology. Higher communism, a political ideal most Chinese political thinkers and leaders have recently advocated, prescribes a hierarchical relationship between humans and machines. Higher communism is a "[…] a society where machines do the labor required to meet our physical and material needs, and we are free to develop as we see fit" (Bell and Wang 2020, 182-183). To achieve the state of higher communism, the given society needs to go through advanced forms of capitalism, in which the productive forces (including technology and the knowledge to make use of it) will need to be so well developed and highly efficient that capitalists can produce a large material surplus without which communism would be infeasible. In this sense AI technologies will be more efficient than any historic forms of technologies in developing efficient means to produce goods and services. Nevertheless, we still need to keep in mind that AI technologies are developed simply because we as humans want to use them to produce a large

material surplus that will liberate us from day-to-day, burdensome work which is arguably often required in a capitalist society.

Bell and Wang further argue that capitalists "[…] cannot just rely on exploitation of labor to squeeze profits" and they are naturally motivated to "[…] constantly revolutionize the means of production to maintain a competitive edge against other capitalists" (2020, 184). Nevertheless, Western humanists such as Nick Bostrom (2016) feel worried that human efforts to continuously develop intelligent entities will eventually lead to a 'singularity' catastrophe in which these intelligent entities are smart enough to turn against humans and treat humans as their slaves and use humans as necessary resources to achieve their own goals (which are different from human goals). Bell and Wang argue that at least in the short to medium term the development of AI will only lead to ethical dilemmas rather than singularity catastrophes. In the long term, they suggest that "[…] a strong and capable pro-human organization that represents the interests of the large majority of people" is needed to ensure that humans maintain their dominance over machines (2020, 186). Political theorist Langdon Winner (1980) famously argues that some inherently political technologies appear to require or to be strongly compatible with particular kinds of political relationships. To some extent, Winner would probably agree with Bell and Wang that more centralized or even authoritarian political regimes are needed to ensure that AI technologies will not turn against us but will only serve human needs.

When discussing the moral standing of AI entities such as robots, similar to Bell and Wang, Joanna J. Bryson (2010) provided a rather controversial position and argued unequivocally that robots should be slaves. Bryson is aware of the negative connotations of slaves in history and makes it clear that she is not at all supportive of the dehumanization dimension of slavery. What she tries to do is to use a radical and extensively controversial concept to alert people that we should not humanize robots *in the first place*. In contrast to racism that has been widely condemned in current society, what Bryson suggests is not that "robots

should be people you own" but that "robots should be servants you own." Therefore, robots should not be described as persons or be provided with legal or moral responsibility for their actions. Humans have the moral obligation to determine the goals and behaviour of robots and thus humans fully own robots. Similar to Bell and Wang, Bryson argues that robots are valuable simply because they have "[…] the potential to extend our own abilities and to address our own goals" (2010, 63). Bryson also warns people that humanizing robots can potentially lead to the dehumanizing of real people and poor human decision making in the allocation of resources and responsibility. For instance, Bryson feels concerned that overly humanizing or anthropomorphizing robots sometimes can waste our time, money, and other finite sources being given to robots that otherwise would be spent on other humans and human-human interaction.

Nevertheless, it is worth examining some potential challenges with Bell and Wang's approach. Despite the fact that Bell and Wang argue that technologies such as AI systems should only be treated merely as tools for human needs, there are circumstances in which anthropomorphism is not only possible but also *necessary* for robots to complete their assigned tasks. In these cases, humans may have the natural tendency to anthropomorphize robots and these robots may appear to have some kinds of moral agency or human-like rights (Coeckelbergh 2010). For instance, a paediatric robot working with autistic children is often designed to be anthropomorphic and it can often be perceived as having some kind of human-like agency. The relationship between this robot and an autistic child shapes how the robot is perceived. It is very likely that at least certain people – such as paediatric patients – may potentially assign some kind of agency to anthropomorphic robots and the hierarchical relationship between these robots and humans is not that clear. Parents and healthcare professionals may also feel less obligated to engage in monitoring children's treatment plan as the robot is perceived to be more knowledgeable. In this sense, robots cannot be merely tools anymore but rather

they mediate the relationships between parents, healthcare professionals and paediatric children.

One potential objection to such a criticism from Bell and Wang's perspective might be that there are potential risks associated with anthropomorphizing robots and engineers should thus intentionally disanthropomorphize them so that they will only be treated as inanimate tools. Scholars have also studied that anthropomorphism can further lead people to overtrust what robots can do (Ullrich, Butz, and Diefenbach 2021). Borenstein, Howard, and Wagner (2017) argue that overtrusting issues partly caused by anthropomorphism can bring risks to children, parents, and healthcare professionals. For instance, robots may be perceived as infallible and thus invite parents to defer judgement about their children's wellbeing to robots. Healthcare professionals may be confounded by whether they need to trust themselves or robots. In this circumstance, robots are not treated merely as tools but appear to hold some kind of agency or competence.

Therefore, to mitigate the risks resulting from the anthropomorphism of robots, engineers should intentionally design mechanisms that make limitations with robotic capabilities visible (e.g. allowing robots to selectively fail or refuse to perform certain tasks) or keep humans in the loop (e.g. demanding direct attention from the user). In this regard, such robots are intentionally designed to be merely tools. To some extent, doing so creates another level of discussion on hierarchical relationship between robots and humans such as healthcare professionals. The question is then whether the actions and judgements made by robots or those made by healthcare professionals are perceived to be more valuable.

AI Systems as Technological Mediators

The second approach to a relational ethic for AI based on the just hierarchy theory argues that AI technologies can serve as mediators that (pro-)actively shape human relationships and help us live our communal

roles well (many of these role relationships are hierarchical). In addition to the moral standing of AI systems, Bell and Wang have further applied the just hierarchy to examine how AI technologies affect the ways in which we develop our role relationships with others. Their work has mainly drawn on resources in Confucian ethics especially Confucian role ethics.

While both deontology and consequentialism place a strong emphasis on the value of following rules (e.g. Kantian categorical imperatives, consequentialist utility principle) for moral actions, Confucian role ethics argues that the key to becoming a good person is to live and reflect on the social roles (e.g. parent, child, citizen, and engineer) one assumes in specific communal contexts. By nature, humans are social, interdependent, and related to each other after birth (Yu and Fan 2007). According to Confucian role ethics, the moral actions we take in different situations are influenced by the specific roles we take in these situations. We as humans all assume different roles which are determined by the relationships we have with others. These different relationships and social roles affect the ways we choose to interact with others. The tone you use to speak to a parent is different from the one you use to communicate with a stranger. The nature of a particular role relationship often evokes feelings and expectations characteristic of that relationship (Ames 2016). Through living and reflecting on these social roles, one gets to cultivate virtues that define the ideal forms of these social roles. Differentiation and fulfilment of these different social roles is critical for a harmonious and flourishing society.

According to Bell and Wang, the development of AI systems can and should be encouraged by our political communities if these technologies help us realize our constitutive commitments or moral obligations prescribed by our social roles (e.g. child, parent). Similarly, technologies that undermine the realization of our constitutive commitments should be restricted. In this sense, technology is never value neutral. Good technologies should always help promote values (e.g. harmony) respected and

maintained in communities. Therefore, the ethics of technology evaluates to what extent and in what ways technology contributes to a process of harmonization. Reliable technological development often leads to "[…] a continuous negotiation and adjustment of relationships between human beings, society and technology" (Wong 2012, 81).

Bell and Wang imagine two paradigm cases involving AI and robots that exemplify how the development of AI technologies can be assessed by Confucian role ethics:

> If an AI-enabled technology can free us from socially necessary work so that we can be easier to spend time caring for our parents with love and compassion, then such technology should be supported.
>
> If a cute-looking robot relives all of our caring obligations and our parents are convinced that the robot truly cares about their well-being, then the parents care more about the robots than their own children. Such robotic technology should be restricted, from the Confucian perspective (2020, 192-193).

In a similar manner, AI-aided teaching should be encouraged only if it helps with learning, but not if it undermines the personal ties between the teacher and the student that underpin their life-long commitments to learning (Bell and Wang 2020). All these scenarios demonstrate that the moral quality of technology is assessed on the basis of *to what extent technology helps us practice our role-based obligations and cultivate virtues for better living these social roles*.

Western approaches to the ethical assessment of technology are often focused on whether technologies invade individualistic and 'universalistic' values such as privacy, liberty, and other forms of human rights that are presented in ethical principles (sometimes in the form of ethics checklists) (Kiran, Oudshoorn and Verbeek 2015). From a Confucian perspective, this approach overlooks an important aspect of moral life: humans are not isolated individuals but rather members of different communities and contributors to different relationships. Their moral decisions are often

shaped by their moral obligations required by these relationships and social expectations about what they ought to do to live their communal roles well.

For instance, from a Confucian perspective, to examine whether autonomous vehicles should be supported is to examine to what extent such a technology helps drivers better nurture their relationships with others especially intimate ones. Autonomous vehicles can be justifiable if the driver can be liberated from driving the car and use the time saved to nurture their relationships with intimate others (e.g. using the time in the autonomous vehicle to videochat with a parent) (Bell and Wang 2020). Scholars have been discussing how autonomous vehicles ought to be programmed to act when these vehicles are involved in some ethical dilemmas on the road (e.g. the trolley problems). One approach to addressing these dilemmas is to allow autonomous vehicles to capture the habit of drivers and let drivers themselves set up ethical preferences for their own autonomous vehicles.

Bell and Wang argue that such a treatment seemingly provides liberty to drivers but in fact fails to consider that drivers of autonomous vehicles will inevitably be influenced by social norms in the community. Drivers can rarely make their own individual decisions as there will often be multiple autonomous vehicles on the road. The relationship between a careful and considerate autonomous vehicle driver and a reckless driver is not equal. They point out that the considerate driver whose autonomous vehicle is programmed to defer to other autonomous vehicles (including those that accommodate reckless drivers) will experience unfairness and therefore change their behaviour for the worse. Instead they suggest that deference (让, *rang*) needs to be reciprocal if designers of autonomous vehicles really want to make deference work. In other words, perhaps deference should be designed as a default option for all autonomous cars. It will be a teaching moment certainly for reckless and less considerate drivers on the road when the deference mode is activated. By building deference as a default mode for all autonomous

vehicles, a moral hierarchy is also being created between considerate drivers and selfish drivers and a truly civil society would certainly value more considerate and civil drivers.

However, it is also valuable to consider some potential objections to Bell and Wang's approach here. Most of these objections in fact target the limitations of Confucian role ethics which is central to Bell and Wang's arguments. Scholars have already explored some of the challenges Confucian role ethics is likely to encounter in practice. Higgins is concerned that the vision of Confucian role ethics regarding family, family roles, and other social roles extended from family roles can sometimes be too idealistic. For instance, she worries that Confucian role ethics may potentially reinforce social roles that are oppressive. Strong demands for conformity to social roles can potentially lead to "[…] oppression of sexual minorities and others who perform their roles atypically" (Higgins 2018, 218). In addition, designing a location tracking app (e.g. Life 360) may seem to be appealing to someone who genuinely cares about the safety of their partner, however such an app may also allow the user to generate or reinforce manipulation and coercion in their relationship. Furthermore, as indicated earlier, Bell and Wang's just hierarchy theory suggests that our relationships with others are hierarchical and therefore our moral obligations toward others diminish in intensity the further we go from those relationships. Therefore, a potential challenge for the application of their theory in AI design is if and how engineers can design AI systems such as robots responsive to hierarchical relationships.

A potential reaction to such a criticism from Bell and Wang's perspective might be that we need to reflect on whether relationships and associated role moralities need to be constrained by moral standards external to these relationships or roles. In a different article, Bell (2018) argues that some more fundamental moral principles are needed to ensure that role-based moral obligations and hierarchical relationships resulting from these roles are morally justifiable. It is also possible that one may sometimes assume multiple roles and moral obligations associated with these roles can

be in conflict with each other. For instance, the relationship between the doctor and the patient may not *per se* be sufficient for designing a caregiver robot whose job includes reminding the patient to take their medication. It is unclear what the robot caregiver is supposed to do when the patient refuses to take medication (e.g. whether the robot needs to prioritize its role in improving human wellbeing or its role in respecting human agency). Therefore, there needs to be some more fundamental moral principles that can help the robot make more realistic decisions in these moral dilemmas.

Regarding the challenge of how to design robots responsive to hierarchical relationships, one potential reaction might be that we need to return to the critical imagination about the constitutive roles robots assume in communities. A companion robot designed to take care of senior citizens has more moral obligations toward the elderly person it serves than other elderly persons. Given that such a robot has more access to the data regarding this particular elderly person's daily routine, it is easier and safer for this robot to make reliable judgment for this particular elderly person than other elderly persons. Similarly, daily interactions between the robot and the elderly person will help the robot better develop virtuous tendencies and learn how to assume the companion role more efficiently in this particular home environment. It will be dangerous and less efficient to allow the robot to treat all elderly persons the same and make the same judgment for all of them. The robot could certainly have some limited obligations toward a strangers in certain predictable and urgent circumstances such as calling emergency services when the is sees a stranger falling on the street. Nevertheless, this robot apparently has more moral obligations to its owner than to the stranger.

IV. Implications for Designing AI Systems

In summary, this paper conceptualizes two approaches to a relational ethic for AI by leveraging Bell and Wang's (2020) just hierarchy theory. On the one hand, AI systems should be merely treated as tools and

designed with inferior moral standing as compared to humans. On the other hand, AI systems are not value neutral and can serve as mediators that help us live our communal roles well and nurture our relationships with others especially intimate ones. This section will briefly discuss some potential implications of these two approaches for designing AI-enabled technologies.

Humans have a natural tendency to anthropomorphize AI systems such as robots. Such a tendency often leads to extensive discussions on the moral standing of robots. It then raises the question whether robots deserve some rights comparable to human rights and how we live with them appropriately. Nyholm (2020) points out that there are at least three reasons why humans tend to anthropomorphize robots and assign agency to them: (i) robots are created to perform tasks within certain domains (otherwise performed by humans, e.g., self-driving cars, care of the elderly); (ii) science fiction has invited us to view robots as agents; and (iii) humans have a history of projecting agency onto different aspects of the world (e.g. saying the sea is 'angry'). While there are cases in which anthropomorphizing is necessary, there are also cases in which it can be unnecessary or even problematic to anthropomorphize robots. The criterion here is to examine whether anthropomorphizing or 'human likeness' is essential for robots to complete their assigned tasks and generate efficient human-robot interaction. It is probably less necessary or even counterproductive to anthropomorphize robots that help you clean your house or work at the assembly line. These robots are simply tools.

Anthropomorphizing strategies are more designing human alike appearances. Darling (2017) uses the concept of 'framing' to describe a set of much broader strategies and tools for influencing how robots are related to us. Framing tools can include the voice, gender, race, background story, or any other forms of narrative we assign to them. By using different framing strategies, robots can be framed as either tools or companions. Robot designers need to ensure that humans do not fall into the 'android fallacy', that is, "[…] the trap of anthropomorphizing

robots and framing them as social agents rather than tools" (Darling 2017, 181). Soldiers and military professionals sometimes develop intimate relationships with robots serving in the military such as land mines detecting robots. There have been stories in which soldiers would risk their own lives to save their robots. Therefore, framing techniques are needed to frame such robots as mere tools rather than companions. Potentially strategies may include avoiding assigning human or animal names to these robots and intentionally referring to them as machines.

AI systems such as robots are not merely tools and they can also serve as technological mediators. They can potentially help us fulfil the moral obligations prescribed by the roles we assume in various communal contexts and thus nurture flourishing relationships with others. Thus, AI designers need first to reflect on what certain ideal relationships mean in particular cultural contexts and such a reflection then becomes a major *source of inspiration* for designing AI systems or robots to better support the flourishing of these relationships. For instance, engineers need to creatively imagine and critically reflect on what an ideal and healthy relationship between an aging parent and their adult child means in a particular context and then design a robot to enhance the moral relationship between the two. Any companion robots that can potentially harm the fulfilment of the role-based moral obligations of either the parent or the child would not be called good design.

In summary, the present contribution conceptualizes a relational ethic for AI based on Bell and Wang's theory of just hierarchy. While engaging the theory of just hierarchy, I also briefly exemplified potential criticisms of such a theory and discussed how Bell and Wang may potentially react to these criticisms. I argue that such critical engagement of scholarship can (i) generate a more comprehensive and deeper understanding of the theory; and (ii) allow philosophical theories such as just hierarchy to be translated into tangible frameworks that can be used to inform practice such as human-AI interaction design.

Works Cited

Ames, Roger T. 2016. "Theorizing 'Person' in Confucian Ethics: A Good Place to Start." *Sungkyun Journal of East Asian Studies* 16/2: 141-162.

Bell, Daniel A. 2018. "Roles, Community, and Morality: Comment on Confucian Role Ethics." In *Appreciating the Chinese Difference: Engaging Roger T. Ames on Methods, Issues, and Roles*. Edited by Jim Behuniak, 203-212. Albany, NY: SUNY Press.

Bell, Daniel A., and Pei Wang. 2020. *Just Hiearchy: Why Social Hierarchies Matter in China and the Rest of the World*. Princeton, NJ: Princeton University Press.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Polity.

Borenstein, Jason, Ayanna Howard, and Alan R. Wagner. 2017. "Pediatric Robotics and Ethics: The Robot is Ready to See You Now, But Should it be Trusted?" In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Edited by Patrick Lin, Keith Abney, and Ryan Jenkins, 127-141. New York: Oxford University Press.

Bostrom, Nick. 2016. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bryson, Joanna J. 2010. "Robots Should Be Slaves." In *Close Engagement with Artificial Companions: Key Social, Psychological, Ethical and Design Issue*. Edited by Yorick Wilks, 63-74. Amsterdam: John Benjamins.

Coeckelbergh, Mark. 2010. "Robot Rights? Towards a Social-Relational Justification of Moral Consideration." *Ethics and Information Technology* 12: 209-221.

Cottine, Cheryl. 2020. "That's What Friends Are For: A Confucian Perspective on the Moral Significance of Friendship." In *Perspectives in Role Ethics: Virtues, Reasons, and Obligation*. Edited by Tim Dare and Christine Swanton, 123-142. New York: Routledge.

Darling, Kate. 2017. "'Who's Johnny?' Anthropomophic Framing in Human-Robot Interaction, Integration, and Policy." In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, 173-188. New York: Oxford University Press.

Hagendorff, Thilo. 2022. "A Virtue-Based Framework to Support Putting AI Ethics Into Practice." *Philosophy and Technology* 35/:1-24.

Kiran, Asle H., Nelly Oudshoorn, and Peter-Paul Verbeek. 2015. "Beyond Checklists: Toward an Ethical-Constructive Technology Assessment." *Journal of Responsible Innovation* 2/1: 5-19.

Klincewicz, Michal. 2017. "Challenges to Engineering Moral Reasoners: Time and Context." In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Edited by Patrick Lin, Keith Abney and Ryan Jenkins, 244-259. New York: Oxford University Press.

Nyholm, Sven. 2020. *Humans and Robots: Ethics, Agency, and Anthropomorphism*. Lanham, MD: Rowman & Littlefield.

Ullrich, Daniel, Andreas Butz, and Sarah Diefenbach. 2021. "The Development of Overtrust: An Empirical Simulation and Psychological Analysis in the Context of Human–Robot Interaction." *Frontiers in Robotics and AI* 8: 554-578.

Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York: Oxford University Press.

Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus* 109/1: 121-136.

Wong, Pak-Hang. 2012. "Dao, Harmony and Personhood: Towards a Confucian Ethics of Technology." *Philosophy & Technology* 25: 67-86.

Yu, Erika, and Ruiping Fan. 2007. "A Confucian View of Personhood and Bioethics." *Journal of Bioethical Inquiry* 4/3: 171-179.