# EuropeNow

About ⌄   Issues ⌄   Perspectives ⌄   Reviews ⌄   Literary ⌄   Visual ⌄   Campus ⌄   The Daily   Partners ⌄   Donate

**RESEARCH**

# Re-thinking "Human-centric" AI: An Introduction to Posthumanist Critique



This is part of our special feature, **Rethinking the Human in a Multispecies World.**

**The problem with "human-centered AI"**

AI is everywhere and nowhere: everyone is talking about it—and some are even studying and developing it—but there is still no operating system like "Samantha" (*Her* 2013) or androids like "Ava" (*Ex Machina* 2015), although we do have Sophia the Robot[1] (*Hanson Robotics* 2016). AIs are being implemented in every sector and domain of society from warfare, medicine, and health, to entertainment and dating; but artificial intelligence research has focused on *narrow* rather than *general* notions of AI (Goertzel 2014). Nonetheless, many are calling for the preservation of "human-centeredness" and anthropocentric principles when dealing with the development and implementation of advanced emergent machine-driven technologies (Fjeld et al. 2020). Leading academic institutions such as Stanford University's Institute for Human-Centered Artificial Intelligence (HAI) propose to align AI development with the notion of "beneficial intelligence" based on humanistic and human-centric values. Generally, human-centrism posits that humans should have command over nonhuman entities. Theoretically, posthumanism offers a substantial basis for challenging this claim. Indeed, problematizing AI in light of posthumanistic critiques would be beneficial in confronting questions of whether nonhuman intelligences can be conceptualized in terms *other than humanistic*. I here aim to re-think the humanistic sense of the "human" and its instrumentalized view of the "nonhuman" as a tool of human ends. Posthumanism is the critical interrogation of the conceptual and experiential boundaries that claim to distinguish the "human" realm from the "non-human" or "other-than-human" (Bellacasa 2017). I suggest that it provides critical and speculative alternatives for thinking about AI.

**Humanism and human-centrism**

A hallmark of humanism is that it established humanity's separate and exceptional character and, purposely or not, led to the subjection of everything else to this alleged special status. Strongly anthropocentric, humanism posits a theory of "human nature" that is used as a basis for making various normative, moral, cultural, and legal claims that elevate humans to the status of moral and political agents while relegating nonhumans to a lesser more instrumental status. Humanism grounds its ethical claims in the human capacities for reason, autonomy, impartiality, and universality, which are then used as justifications for mastery and management of nonhumans who are considered to lack these capabilities. In the intellectual histories of Western thought, the view that humans possess unique capacities that make them exceptional and/or superior to others is often found. For instance, ancient Greek virtue-ethics, medieval humanism, early modern

mechanism, and even contemporary philosophy of mind are grounded in anthropocentric terms that privilege the achievement of human ends by way of human rationality at the expense of nonhuman lives.

In human-centric narratives, such as the Renaissance notion of "man as the measure of all things," or in modern debates about "nature versus nurture," or in contemporary debates about "humans versus robots," humans are framed as having special insights and being self-authorized to preside, command, and control all others. Take for example Aristotle's stance justifying slavery, which is dependent on connecting various hierarchies together, namely the human domination of nature, male domination over females, the master's domination over the slave, and Reason's domination of the body and emotions. Aristotle offers an account of each dualism's place in a chain of hierarchies:

It is clear that the rule of the soul over the body, and of the mind and the rational element over the passionate, is natural and expedient; whereas the equality of the two or the rule of the inferior is always hurtful. The same holds good for animals in relation to men; for tame animals have a better nature than wild, and all tame animals are better off when they are ruled by man; for then they are preserved (quoted in Plumwood 1993, 46).

Philosophically and politically, this conceptual network of binaries—mind/body, reason/emotions, human/animal, male/female, freedom/slavery—reflects the dynamics of mastery and/or hierarchy: of higher over lower, superior over inferior, essential over instrumental. Historically, humanism has portrayed the "human" as agent, as creator of culture and technologies, and as a bearer of rights and responsibilities that makes use of other life forms (both non-human and human) including animals, plants, and machines. Those who have been regarded as deficient in rationality and intrinsic moral worth such as women, children, slaves, and colonial subjects are deemed as lacking full human potential and are treated as less than human, even as "sub-human." The instrumentalization of techniques and technological manipulation become the main vehicle by which humanism perpetuates this master(y)-logic. Thus, many argue that such human-centrism follows a "logic of colonization" (Plumwood 1993), conceptualized dualistically as the mastery of a superior (order) over an inferior (one): "[t]he upperside is an end in itself, but the underside has no such intrinsic value, is not for-itself but merely useful, a resource" (ibid, 53). "This is a model of domination and transcendence in which freedom and virtue are construed in terms of control over, and distance from, the sphere of nature" (ibid, 23). Within humanism, humans are depicted as capable of transcending their animal roots through intellection and instrumentalization of a nonhuman order for the benefit of humankind. Liberal normative theories of human rights are grounded in this human-centric representation of the individual who is expected to take ownership over its own self, this self-mastery thereby sanctioning the exercise of mastery over others who are incapable of such self-legislation. In the history of Western ethics, such human exceptionalism has centered on the human intellect, especially the activity of deliberating about human ends, which require mental and practical capacities to discern the worthy ends of human life.

By means of dualism, the colonised are appropriated, incorporated, into the selfhood and culture of the master, which forms their identity. The dominant conception of the human/nature relation in the west has features corresponding to this logical structure (ibid 1993, 41–42).

Within this mastery model, humans are supposed to govern unpredictability through the instrumentalization of their rationality and their normative and norm-making capacities. Strong human-centrism posits the achievement of human control using the instruments of reason (like animals and machines). This image of human control—of the morally conscious, modern individual that technologically transforms the nonhuman world for the benefit of all—is such a pervasive but unquestioned dogma that to challenge this viewpoint amounts to disrupting prevailing ways of doing, thinking, and being. Even critically minded liberal thinkers like Mary Wollstonecraft did not challenge the human-centric presumption regarding "man's pre-eminence over the brute creation:" "the perfection of our nature and capability of happiness must be estimated by the degree of reason, virtue and humanity that distinguish the individual and from which the exercise of reason, knowledge and virtue naturally flow" (quoted in Plumwood 1993, 17).

Historically, advocacy for the rights and welfare of those deemed to lack reason (and thus considered non-rational) did emerge among liberal sentimentalists such as Jeremy Bentham, who argued that non-rational people should be protected *not* on the basis of rational capacities and claims to freedom and equality, but rather because the "non-rational" have shared capacities for sentience. They are therefore owed limited human protection and sympathy. Liberal sentimentalism sought to protect individual freedom by borrowing from nineteenth and twentieth-century ideals of social equality as minimal capabilities that the state must guarantee and should also extend to nonhuman animals, people with disabilities, and noncitizens (e.g., Engster 2006; Nussbaum 2007).

These revisions, however, do not overturn the unchallenged assumption that what makes nonhumans worthy of moral consideration is their commonality, similitude, and resemblance with humans who have a special status as "moral agents." Thus, liberal concepts of human moral agency, even when they go beyond possessive individualism, tend to assess the worth of nonhumans in terms of human-centric standards. As Willett argues, displays of vulnerability and appeals for sympathy do not suffice to generate the solidarity that an egalitarian political ethics requires (Willett 2014, 38).

While most discussions of AI are still anchored in humanistic/humancentric narratives, there are reasons for rejecting this model and for turning to alternative worldviews. The

notion of "nature" is a complex and contested battlefield of meanings, hierarchies, and exclusions where racial, sexual, ethnic, and other differences have been cast in terms that distinguish so-called "higher" forms of humanity from "lesser" ones deemed to lack some degree of rationality or cultivation. The master/slave dichotomy at the heart of humanistic version of human control reproduces a cluster of other dualisms such as self/other, culture/nature, human/animal, human/machine, man/woman, colonizer/colonized etc. This logic of mastery/subjugation views domination as natural and befitting. Within this model of control, "the multiple, complex cultural identity of the master [is] formed in the context of class, race, species and gender domination"; the problem, however, is that "the assumptions in the master model are not seen as such, because this model is taken for granted as simply a human model" (Plumwood 1993, 5, 22).

Discussions of AI have tended to focus on the prioritization of human-centered epistemologies that conceptualize humans as being at the center of agency, cognition, and broader relations/networks of exchange. In philosophical discussions, emphasis remains on the moral problems that might occur with increasingly intelligent artificial machines and rebellious robots, and whether robots can be designed to act morally and serve human needs (see for example, Boddington 2020; Gunkel 2018). Notwithstanding the diversity of differing ethical viewpoints, the legitimacy of human-centrism goes largely unquestioned, as does the possibility of imagining alternatives. Thus, "human-centered AI" tends to be strongly anthropocentric. As Boddington (2020) asserts, humans have agency —especially moral agency—which is an essential attribute of humanity and what makes it distinct and valuable: "This we must not lose. Computers, even those with artificial intelligence, are our tools. They should not diminish our agency; ideally, we should use them to enhance our agency.

A dominant portrait of human-centric AI posits techno-scientific regulation as key to a beneficent future in which the threats and risks of artificial intelligence are managed by normative constructions of control and containment. While contemporary proposals for "human-centered AI" envision governance controlled by public institutions that promote and support human flourishing and liberal ideals, such strongly anthropocentric responses reinforce the principle of human mastery over nonhumans. In this sense, the master model of strong anthropocentrism replays a leitmotif of modernity: the transformation of chaos into order through human ingenuity and control (Hurlbut 2018, 147). By deploying this logic of mastery, framing AI in terms of human moral agency fails to question the logic of domination and transcendence that defines prevalent conceptions of human/nonhuman relations.

### Critical posthumanism

The rationale for an alternative to humanism emerges when considering the limitations of the mastery/subjugation model. Critical posthumanism seeks to deprioritize and weaken human-centrism, rejecting individualism, and instead underscoring the compatibilities between human animals, nonhuman animals, and machines. Cynthia Willett's (2014) concept of "interspecies ethics" highlights the limitations of the liberal model of human agency and offers a posthumanist lens that goes beyond "modern and postmodern binaries ... to engage multilayered symbiotic agencies and biosocial communities" (7). Critical of the legacies of humanism, critical posthumanists place humans "on the loop" with nonhumans, emphasizing co-evolution between humans and nonhumans and thus attempting to weaken anthropocentrism. Humans and nonhumans are conceptualized as co-producers under specific but changing environmental conditions. Criticizing the scientific imagery that segregates species and privileges human-centric forms of life, critical posthumanists argue for a rejection of the principle of human mastery in favor of a conceptualization that bridges the divide between human and nonhuman lives. The emphasis on posthumanism is intended to devalue strong anthropocentrism as a way of overturning the humanistic narrative of control: whereas the mastery model makes humans the creator and controller of technological change, critical posthumanism would claim that the "human" is an open-ended category and the product of ongoing processes of collective bio-socio-technical individuation. For instance, Rosi Braidotti (2013) has argued that life is not the right of the human species alone; rather, it is the very force that connects various species, and also cuts across them. Life is a "transversal alliance across species and among posthuman subjects" that "opens up unexpected possibilities for the recomposition of communities, for the very idea of humanity and for ethical forms of belonging" (60, 71–72).

Thinking of AI in posthumanistic terms is still relatively absent from the current discussions. One conceptual challenge may be that critical posthumanism does not entail ceding human oversight altogether; the anthropocentrism of this approach, however weak, remains. Although anthropocentric thinking may not be suitable for imagining futures that include humans but that do not privilege them in any way, it is especially difficult to overturn completely when the intellectual resources available are so embedded in human-centric histories. Both humanism and critical posthumanism retain normative assumptions about why humans should expect to maintain a special status among nonhumans.

### Speculative posthumanism

Conceptualizing non-anthropocentrism may require going beyond conventional ways of thinking. The task of conceptualizing non-anthropocentrism would thus fall to a distinct/third kind of posthumanism, which I call "speculative," in which human control would be deprioritized and nonhuman rationales would be prioritized. Since there are few

ontological and epistemological resources that are not somehow connected to human-centrism, the conceptual task would require speculative rather than normative thinking. The question of governing AI in the specter of non-anthropocentrism would have to address the question of how humans will conceive of being governed by AIs that have evolved beyond the capacities and powers of the human species. Speculative posthumanism would displace humans from the seat of command and entail the phasing out of humans and human-centered perspectives altogether. In speculative posthumanism, anthropocentrism is devalued, deposed and eventually jettisoned; control would be disconnected from any human-centered values. Both strong and weak anthropocentrism would be displaced by nonhuman governmentalities that would go beyond the horizon of the human.

Speculative posthumanism speculates about alien/*xeno*-intelligences well beyond human parameters. Accordingly, speculative posthumanism argues for a "disconnection thesis," the idea that humans should not be conceptualized in terms of the presence or absence of some essential "human" of personhood, but as "an emergent disconnection between individuals [that] should not be conceived in narrow biological terms but in 'wide' terms permitting biological, cultural and technological relations of descent between human and post-human" (Roden 2015, 105). Instead of positing any anthropocentric baseline (not even a weakly constrained one), speculative posthumanism would begin with the assumption that "our current technical practice could precipitate a nonhuman world that we cannot yet understand, in which 'our' values may have no place" (Roden 2015, 124). Imagining AI/human relationalities through speculative posthumanism may allow alternative conceptions of humanity in which human-centrism is marginal. Here, "human" would not refer primarily to the humanistic picture associated with biology or cognition, but to a view that is disconnected from any human-centrism and quite alien.

### Beyond humanism: critically confronting AI

Strongly anthropocentric responses to framing "ethical AI" need to be challenged. Proposed frameworks should explore the range of different perspectives beyond humanism. As many feminist and post-colonial scholars have warned, (post)modernity has found toeholds in a variety of optimistic futures tied to neoliberalism, the most popular perhaps being the narrative of transhumanism or extropianism, the assimilation or supersession of the human in the suprahuman machine. Posthumanism takes a critical view of this scenario, interrogating it for its triumphalist rupture from the animal, its complicity with the class politics of big capital and its fantasmatic investment in patriarchy (Banerji and Paranjape 2016, 2).

The uncritical humanistic drive underlying current proposals for human-centered AI would suggest that current AI ethics will be limited and might even reproduce the values that have sanctioned social, political, and economic hierarchy, exclusion, and subjugation to date. In this regard, the importance of interdisciplinary/intercultural dialogue and debate cannot be overestimated. More critical and speculative posthumanistic conceptions may decenter strong anthropocentrism by upholding relationality, solidarity, and care as primary aspects of human/nonhuman associations (rather than atomism, hierarchy, and mastery). As I have tried to show, theories of posthumanism are compelling for potentially pushing conceptual boundaries beyond anthropocentrism.

***Nandita Biswas Mellamphy*** *is Associate Professor of Political Science; Affiliate member in Gender and Sexuality Studies; Core faculty in the Centre for the Study of Theory and Criticism; and Current Director of The Electro-Governance research group at Western University in Canada.*

## References

Banerji, Debashish, and Makarand R. Paranjape. 2016 "The Critical Turn in Posthumanism and Postcolonial Interventions." In *Critical Posthumanism and Planetary Futures*, ed. Debashish Banerji and Makarand Paranjape, 1–12. New Delhi: Springer India.

Bellacasa, Maria Puig de la. 2017. *Matters of Care*. Minneapolis: University of Minnesota Press.

Boddington, Paula. 2020. "AI and Moral Thinking: How Can We Live Well with Machines to Enhance Our Moral Agency?" *AI and Ethics*. https://doi.org/10.1007/s43681-020-00017-0.

Braidotti, Rosi. 2013. *The Posthuman*. Cambridge: Polity.

Engster, Daniel. 2006. "Care Ethics and Animal Welfare." *Journal of Social Philosophy* 37 (4): 5236.

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. 2020. "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI." 15 January. Berkman Klein Center Research Publication, No. 2020–1. http://dx.doi.org/10.2139/ssrn.3518482

Goertzel, Ben. 2014. "Artificial General Intelligence: Concept, State of the Art, and Future Prospects." *Journal of Artificial General Intelligence* 5(1) 1-46. DOI: 10.2478/jagi-2014-0001

Gunkel, David J. 2018. *Robot Rights*. Cambridge, MA: MIT Press.

Hurlbut, J. Benjamin. 2018. "Remembering the Future: Science, Law, and the Legacy of

Asilomar." In *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, ed. Sheila Jasanoff and Sang-Hyun Kim, 126–151. Chicago: University of Chicago Press.

Nussbaum, Martha. 2007. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge: Harvard University Press.

Plumwood, Val. 1993. *Feminism and the Mastery of Nature*. London: Routledge.

Roden, David. 2015. *Post-human Life: Philosophy at the Edge of the Human*. London: Routledge.

Willett, Cynthia. 2014. *Interspecies Ethics*. New York: Columbia University Press.

[1] https://www.hansonrobotics.com/sophia/

Photo: Calne Wiltshire UK May 22 2020 The Head, a metal sculpture of a head, by Rick Kirby, winner of a competition to celebrate the millennium in the Wiltshire town of Calne | Shutterstock

Published on November 9, 2021

**Share:**

EuropeNow: a journal of research & art