

Review of: "[Essay] Not Quite Like Us? — Can Cyborgs and Intelligent Machines Be Natural Persons as a Matter of Law?"

Debora Halbert¹

¹ University of Hawaii System

Potential competing interests: No potential competing interests to declare.

The foundation of a good contribution to the scholarship is when an essay makes you think. Reading Professor Gervais' piece made me want to engage in a conversation throughout about the issues raised and offer other possible perspectives. While the piece raises thought-provoking ideas, I don't necessarily agree with the stated goal – to find a distinction between humans and machines that the courts could operationalize. That being said, a legally operational definition would have both merit and usefulness if for no other reason than to demonstrate to whom we will extend rights in the posthuman future. My comments are primarily about how I engaged with (and found engaging) the essay.

Establishing criteria for how AI will acquire (or be denied) legal personhood is an important subject of discussion. It is also important to have this discussion prior to the full emergence of the type of AI that may require such definitions. However, by the end of the essay, I remain unconvinced there is a clear distinction to be made. Attempts to ground a distinction in moral philosophy or essential human characteristics may work in the short term, but not in the face of the posthuman reality that faces us. If, as Professor Gervais argues, "a line can be drawn to separate an entity holding human rights by 'nature', and an entity incapable of doing so (p. 16)," I am not sure why the line can and should be drawn there. I question if "nature" is a relevant category given the already existing intersections between humans and technology and I wonder why in the trajectory of the concept of legal personhood, natural personhood is the relevant line as a matter of law. A definition of legal personhood and who can be extended rights under the law needs to acknowledge the role power plays in making such a determination and what interests such a definition is designed to protect or enable.

Professor Gervais begins with a discussion of sentience, sapience, and intelligence as relevant concepts for understanding the human versus the nonhuman. He settles on sapience as the factor that can be used to distinguish humans from machines (p. 4). Sapience is defined as the way reason and emotion combine to guide our behavior within the mind and body (p. 4). Central to this essay is the *embodied* nature of human sapience as distinguished from current AI that is digital rather than biological. Gervais acknowledges that AI could acquire a level of sapience, and so from the start the usefulness of drawing the line here is questionable. Afterall, Kurtzweil's claims regarding the singularity and efforts to achieve it could mean that in the no-so-distant future AI will be biologically embodied. How do we distinguish then?

Gervais suggests that perhaps one can delineate between humans and machines using human language. For example, he suggests that humans are the only ones "fully able to inductively understand and use human language (p. 11)." I am

not sure I find the claim that, “machines can increasingly use human language but that is not how they understand the world (p. 12)” as a compelling distinction either. Certainly, the coding for an AI is binary, but the philosophical question posed by machine learning is when does the manifestation of this code become more than the sum of its parts? I will of course defer to Noam Chomsky on all matters linguistic but I think more attention to why language is the distinction that matters is needed. Additionally, given its ability to use human language, how might this delineation be operationalized for the purposes of the law? I’d like to know more.

As I understand it, the ultimate goal of the paper is to create a legal line to distinguish between AI and humans to avoid extending them legal personhood or to clarify when they should be extended legal personhood. However, even as the paper goes through the different dimensions of sentience, sapience, moral philosophy, the use of reason, and ethics, there is no naturally occurring bright line where such a distinction can be drawn. Even the biological basis used for the adoption of sapience as a starting point can be challenged. Gervais references the literature on cyborgs, but doesn’t push far enough because the merging of humans and machines will indeed provide challenges that cannot use embodiment as the criteria for understanding what is uniquely human or posthuman.

Cited in the references are Clark and Chomsky’s extended mind philosophy, but their thesis doesn’t really make it into the paper regarding how fundamentally technology may alter the interface between the biological and the technological. Being self-referential for a moment, I’d suggest looking at *The Neuropolitics of Brain Science: and its Implication for Human Enhancement and Intellectual Property Law* (2020) <https://www.mdpi.com/2409-9287/5/4/33> as another effort to engage in theories of the extended mind, contemporary brain science, and its implications for the law.

Ultimately, each effort to make humans unique can be undermined by either the application of the same arguments to the animal world, the natural world, or to current technologies. Locating humanness in biology and the embodied nature of reason and emotion may work for the time being, but even this definition later in the paper is challenged when questions about when a human stops being human as new technological enhancements alter the human body’s relationship to cognitive and physiological support systems (p. 18). To reference the singularity again, the distinctions being drawn here will most likely be blown away as AI, nanotechnology, and biotechnology produce the posthuman.

What I would find useful and engaging is the focus on the concept of legal personhood, which is the relevant kind of “person” the law allows to have standing. As Gervais notes, natural personhood is to be distinguished from legal personhood. Perhaps establishing that the concept of legal personhood is premised upon hegemonic discourses used to establish control through the law would be an avenue towards a new posthuman definition. After all, the patriarchal systems of the past are the reason women were not granted legal personhood until the 19th century (and even later depending on how you want to define it) and capitalist hegemony in the United States is the reason why corporations can be legal persons despite being unnatural. Knowing who benefits and structures the law to privilege access and control is key to understanding the issues. It is why some states have established legal personhood for fetal cells or why Judge Staton’s makes an argument for the standing of future generations in her in *Juliana v. U.S.* (2020).

In other words, perhaps instead of attempting to ground the idea of legal personhood in a rational and moral foundation, this essay could take the stand that there is nothing about humanity making them unique. Our legal distinctions about who

can (and cannot) have access to the law is based upon hierarchies' judges and lawyers defend as society evolves. It may be that what makes humans unique is their constant need to loudly proclaim their uniqueness and then impose their presumed superiority on the other – other humans, other creatures, and ultimately upon their own creations.

So, if an argument needs to be made that AI should not be given the status of a legal person, there seems to be no intrinsic reason to locate that argument in something essential to humans, but rather perhaps to use the logic of corporate personhood – corporations follow their own moral codes, are disembodied entities, are not self-aware, yet because they are extensions of humans and serve the interests of a capitalist society, have been granted legal personhood and constitutional rights in the United States.

The essay offers a justification for why corporations should not be used as the relevant metaphor, but I think more is needed here (p. 19). If a corporation exists to shield their human creators from the consequences of their (the human's) moral lapses, then why would we extend them legal personhood under the definitions offered in this essay when that definition is grounded in an embodied morality and reason? If corporations are an extension of their human's agency and can act immorally, then why use morality grounded in embodied human nature as a foundational principle for AIs? How is the argument that a corporation is a "mere legal construct" a refutation when legal personhood itself is also such a construct (p. 19)? In the case of an AI, the question then becomes why would we *not* want them to have legal personhood?

One reason for the claim that an entity must be embodied in order to have legal rights (minus corporations of course) is the challenge of how one would punish a disembodied entity. However, as the essay points out in footnote 42, there are examples where ships are punished, or other nonhuman entities had legal personhood so that humans could exact vengeance (p. 24). I'd suggest, rather than a footnote arguing for the exception, that this idea becomes the claim for demonstrating that any distinction is designed to benefit humans who control the law and its outcomes. Even embodiment may not be critical if one can punish a disembodied nonentity.

Throughout the essay, I wrote notes to myself regarding how I might respond and/or think about the effort to operationalize the distinction between the human and non-human. Most central to my inquiry was trying to determine if the goal was to find the line between humans and machines and defend it or to show that all lines are arbitrary. In my reading, even as the essay seems to want to do the former, it ends up clarifying the latter. The fact a corporation could be a person and a woman could not, suggests that sapience isn't a particularly relevant framework for legal definitions. If the claim is that the distinction between humans and machines must be biological and that moral agency is rooted in biology (p. 17) then the very next section tends to undermine this claim by theorizing about the role cyborgs play in altering the human (p. 18).

So, part of the question is: are we establishing this definition for today or for the possibilities of this technology in the not-so-distant future? Is the goal to assure that AI will never achieve the legal status of humans? Or to design the conditions under which AI would be held legally culpable? In today's world, people have acquired AI girlfriends, therapists, and priests, albeit very simple and not sentient apps so far. Because we do not see these as sapient (or sentient) they do not

have the same legal culpability. At what point, however, does that legal culpability shift from the AI designer to an AI? I'm not sure that embodiment is the relevant factor.

The paper concludes that humans should not grant AI legal personality until "their degree of moral agency is sufficient." I think this is an interesting line to draw, especially given the murderous actions of some corporations. It could make sense, though I am still unclear when AI would achieve it. When an AI can be held culpable for an action? When it can AI assert its rights against a human in court? If it does gain a certain moral agency, would an AI have standing to sue? Could the AI be held liable for civic and criminal acts? How might an AI defend itself? How might an AI be punished?

I am left wondering if this is an argument for a philosophy of oppression of AI so that humans will not be replaced but have legally sanctioned control over whatever manifestation of intelligence, autonomy, and personhood might be emergent or if this is this an argument for the legal rights nonhuman entities ought to have? In fact, I think this essay could more clearly articulate what these distinctions will be by including a section that summarizes the discussion into a clearly operationalized definition that the essay is designed to create.

Then, just to push the limits into a more theoretical realm – what makes us even believe we are not ourselves simulations and whatever essence of reality we ground ourselves in is nothing more than a different kind of code? Why not start there?

<https://www.scientificamerican.com/article/confirmed-we-live-in-a-simulation/>. I'd love to have this, and the other conversations the essay inspires over drinks at a conference someday.