ICRES 2018: International Conference on Robot
Ethics and Standards, Troy, NY, 20-21 August 2018.
https://doi.org/10.13180/icres.2018.20-21.08.017

APPROPRIATENESS AND FEASIBILITY OF LEGAL PERSONHOOD FOR AI SYSTEMS

BENDERT ZEVENBERGEN*

Center for Information Technology Policy, Princeton University
310 Sherrerd Hall, Princeton, NJ 08544, U.S.A.
benzevenbergen@princeton.edu

MARK A. FINLAYSON

School of Computing and Information Sciences, Florida International University
11200 S.W. 8th Street, CASE Building, Room 362,
Miami, FL 33199, U.S.A.
markaf@fiu.edu

MASON KORTZ

Harvard Cyberlaw Clinic; Berkman Klein Center for Internet & Society, Harvard University
Wasserstein Hall Suite 5018, Cambridge, MA 02138, U.S.A.
mkortz@cyber.harvard.edu

JANA SCHAICH BORG

Center for Cognitive Neuroscience & Kenan Institute for Ethics, Duke University
140 Science Dr., Durham, NC 27708, U.S.A.
janaschaichborg@gmail.com

TJAŠA ZAPUŠEK

Faculty of Law, University of Copenhagen
Njalsgade 76, 2300 København S, Denmark
tjasa.zapusek@gmail.com

The European Parliament has adopted a proposal to explore the impact of a legal personhood category for AIs, comparable to corporate personhood ("AI Personhood"). We propose that it is premature to introduce AI Personhood, primarily because (i) the scope of AI is still ill-defined, (ii) the potential economic efficiencies and distribution of gains is uncertain, (iii) the ability of existing legal structures to achieve similar ends have not been sufficiently analyzed, (iv) the moral requirements for personhood have not yet been met, and (v) it is not yet possible to assess the social concerns arising from AIs that are indistinguishable from humans. To support our conclusion, we discuss (1) the relevance of legal personhood, (2) the definitional difficulties surrounding AI, (3) currently applicable legal principles, (4) the potential benefits and drawbacks of AI Personhood, and (5) the conditions that might justify such a category in the future. We propose five specific necessary conditions—technological, economic, legal, moral, and social—for AI Personhood, but observe that the conditions have not yet been met and seem unlikely to be met soon.

Keywords: Electronic Personhood; Legal Frameworks; Liability

In February 2017, the European Parliament adopted a non-legislative resolution noting that the increasing autonomy of robots or artificial intelligence (AI) systems raises serious questions as to whether the ordinary approaches to liability are sufficient to ensure just outcomes. The resolution called on the European Commission to explore the liability implications of robots and AIs and, in particular, they raised the possibility of granting AIs status as legal persons ("AI Personhood") [1, § 59f].

*Bendert Zevenbergen is the lead author; the other authors are listed alphabetically.

Granting such "status of electronic persons"—comparable to the legal personhood assigned to corporations—is not a new idea,[2,3] and is a very real legislative possibility. This, combined with the rapid advances in robotics and AI, suggests that it is timely to carefully reconsider the arguments for and against creating a new legal category of AI Personhood.

Our conclusion is that while there may be future conditions that justify or even necessitate AI Personhood, it appears premature and probably inappropriate to introduce AI Personhood now, primarily because (i) the scope of AI is unclear, as a concept or as an artifact, (ii) it almost completely opaque what economic efficiencies will be gained, and what the distribution of economic benefits will be, (iii) we have not demonstrated that existing legal structures cannot achieve similar ends, and (iv) AIs do not yet meet the moral requirements for personhood, and are unlikely to meet them soon.

To support this conclusion, we first discuss the relevance of legal and moral personhood with respect to legal systems generally[a]. We highlight the definitional difficulties of AIs, and the problems these pose for AI Personhood. We next outline existing legal options for addressing the supposedly new problems introduced by AI, and then discuss the advantages and disadvantages of AI Personhood. Finally, we outline conditions that might justify or even necessitate AI Personhood, and conclude that these conditions have not yet been met and are unlikely to be met soon.

1. Relevance of Legal (and Moral) Personhood

Legal personhood is a construct that can be attributed at the will of the legislature, and not necessarily be driven explicitly by moral considerations. On the other hand, the commonsense concept of personhood is tied to being a human. In most legal systems there is a distinction between natural and legal persons: "Natural persons" includes all and only humans, whereas "legal persons" can exclude some humans but include non-human entities that have been deemed as needing special status. For example, societies may count corporations, nations, or political organizations as legal persons (see, e.g., Article 47 of the Treaty on European Union, or[4,5]), whereas some people—such as slaves—were historically denied legal personhood.

A legal person is an entity that can bear rights and duties,[6,7] such as the ability to own property, conclude contracts, or be sued. This is a "legal fiction" that allows non-human entities to be treated like natural persons for some aspects of the law. Without the concept of legal personhood, persons injured or harmed by a faulty product would need to find the exact person responsible within the company producing the product. Legal personhood allows the injured party instead to hold the company responsible. This is often justified by appeal to legal and economic efficiency.

Legal personhood may sometimes be founded on arguments about moral status. This usually means that an entity can suffer, or be able to reason about its own existence and moral responsibilities. Jeremy Bentham and Peter Singer have both argued that moral status is derived partially from our—or an animal's—ability to suffer.[8,9] Suffering can be physical, emotional, or financial. Kant and Regan argued that moral status is derived from intrinsic worth, our sophisticated cognitive capacity to reason about ourselves as "subjects of life".[10,11] Both definitions imply some form on consciousness. beyond the scope of this paper) By way of example, corporations can suffer financially and otherwise, though not exactly like natural persons. Further, a corporation is comprised of people who collectively can reason about the corporation's existence and its moral duties.

[a] A comprehensive review of legal personhood and specific legislation relevant to AIs is beyond our scope here. Therefore, we take a generalized and high-level view not tied to a particular jurisdiction.

2. Definitional Difficulties with AI

The first challenge when evaluating whether AIs should be assigned an existing form of legal personhood, or whether AI Personhood should be created at all, is to define what counts as "AI". The concept of AI itself is notoriously elusive, with different groups strongly disagreeing over precisely what it requires.[12] When John McCarthy coined the term "artificial intelligence" in 1955, he defined it as "the science and engineering of making intelligent machines".[13] Most definitions follow this lead by describing AI as "intelligence exhibited by machines." Common variants add that AI must demonstrate "human" or "human-like" intelligence.[14] Such definitions assume that *intelligence* is clearly defined itself, though it too is ambiguous.

We can further distinguish between narrow and general AI (a.k.a., *artificial general intelligence*, or AGI, for the latter). Narrow AI addresses specific applications, where machines often outperform humans in speed, accuracy, and efficiency. Narrow AI is already widely used. General AI, by contrast, requires intelligent behavior that is (at least) as broad, adaptive, and advanced as a human across a full range of cognitive tasks.[15] It is debatable whether consciousness is a prerequisite for intelligence, or vice versa, and also if and when general AI will be achieved. While we recognize that there is no settled definition of AI, for the purpose of evaluating AI Personhood, we define AI as *human-created digital information technologies and associated hardware that displays intelligent behavior that comes not purely from the programmer, but also through some other means.* We explicitly mention that AIs are man-made, because the designers selected the parameters within which an AI operates and learns. This also acknowledges that AIs include not only software, but also physical hardware.[14] Finally, by mentioning *some other means*, we acknowledge that AIs can develop a type of agency due to the influence of external factors on its behavior.

This definition, then, leads us to the first major problem for AI Personhood, namely: what entity specifically should be accorded the legal status? The hardware and software for an AI system can, in principle, be widely distributed, either physically or computationally.[16] The European Parliament appears to consider AIs as easily identifiable artifacts, even though this is misleading.[17] What about AIs which do not have a precisely defined embodiment, beyond the specific computer on which they temporarily reside? There is no clear answer to this problem at this time.

3. Existing & Current Law

Existing legal approaches treat AIs simply as tools. The European Parliament expressed two primary concerns with this view: that it will be difficult to establish a causal link between the harmful action of the AI and a legal person that can be sued, and that it will be difficult to identify the correct defendant when technologies from several different sources effect an AI system's behaviour. AI Personhood could solve these problems, as the European Parliament suggests, but so could other, less sweeping legal doctrines. We consider a few of those doctrines here.

A party injured by a product with an embedded AI system could hold the producer of the system liable under the doctrine of strict product liability. Under strict product liability, the supplier of a product is liable for harms caused by defects in that product, regardless of whether this was the result of negligence.[18,19] Strict product liability has two features that make it an appealing model for AI liability. First, in many jurisdictions, an injured party can prove that a product is defective without precisely identifying the defect, so long as they can show that the product was less safe than a reasonable consumer would expect and that the malfunction was not due to some external factor; the burden is then on the supplier to disprove or excuse the defect.[20] A system of strict AI liability might create a similar

presumption that any AI system that falls below some pre-defined acceptable rate of error is defective. Second, multiple parties can be joint and severally liable for the same defective product. If one component of larger system is defective, both the component manufacturer and the product assembler can be held liable.[21]

Imagine a person was injured by a defective autonomous vehicle. Under a strict product liability regime, that person could sue the vehicle manufacturer without identifying whether the defect was in a physical or AI component or proving that the defect was the result of negligence. However, the manufacturer would have an incentive to determine whether the defect was in the AI—if so, it could sue the AI developer for indemnification. This system would leave the difficult task of identifying AI defects to parties more capable of doing so than the end consumer and might incentivize development of more transparent AIs. It could also encourage AI developers and product manufacturers to apportion liability preemptively, avoiding the costs associated with *post hoc* litigation.

Product liability is generally limited to physical injuries caused by physical products.[22] How could the legal system handle, for example, a pure software financial AI that loses its clients' money? One possibility is to focus on the decision to deploy the system in the first place. The law could declare that some uses of AI are *ultrahazardous*, meaning they pose a significant risk of harm even when performed with care. Similarly, AIs could be treated as animals. Although animals are autonomous, their owners are legally responsible for them.[23] Under either approach, the law would permit recovery for at least some losses caused by AIs without requiring the injured party to prove that any particular human acted negligently or wrongfully.

Finally, we could consider *vicarious liability* for AIs. Vicarious liability is a doctrine under which one party takes legal responsibility for the conduct of another.[24] If an AI system committed a tortious act, liability would be determined as if the owner had committed that act. The owner could not evade responsibility by claiming lack of knowledge or intent.

4. Advantages of AI Personhood

Even though we have argued that there are ample existing legal mechanisms that could potentially address the issues that AI Personhood is intended to solve, the creation of a new AI Personhood category is nonetheless a real possibility. We see two main potential benefits. First, there are potential instrumental advantages to AI Personhood, which are suggested by analogy to corporate personhood. Corporate personhood allows a connected group of persons—though potentially distributed in time or space—to pool resources and centralize risks. This pooling of resources can be necessary to spur large-scale innovations or to take advantage of economies of scale. In turn, the economy and society in general derives a benefit. From a legal point of view, corporate personhood allows single organizations to be held liable for harms without the need to identify a responsible individual. Legal efficiency is achieved because it allows plaintiffs to sue the organization directly without going through a lengthy, expensive, and arduous process of identifying the specific individuals responsible. Economic efficiency is achieved by the pooling of resources to increase productivity, while creating legal certainty improves the efficiency of operation.

A second benefit is based on morality coupled with a potential technological trajectory: If, in the future, a general AI system is developed that is indistinguishable from a person, by what argument do we deny that system the same rights as a human? More strongly, if an AI can be shown to have real consciousness, suffer real pain, or be truly independent, a majority of the population might feel morally compelled to grant the AI the same rights and responsibilities as humans. While it is clear that AIs are not yet at this level (and it is unclear if or when they will reach it), it would be unwise to dismiss this possibility completely. At

that point, the question may become less an issue of legal or economic efficiencies as it is of "human" rights.

5. Disadvantages of AI Personhood

There are many possible disadvantages of AI Personhood. We list four scenarios that we see as most likely in the near term. First, the European Parliament suggest creating a collective insurance fund to cover damages arising from AIs. However, the technological trajectory of AI is uncertain and unpredictable, and it is therefore unwise to construct financial compensation resources today to meet as yet unknown future needs.

Second, AI Personhood would allow producers and owners of AIs to shift liability to the artifact itself. This will disincentivize investment in adequate testing before deployment. AI Personhood could thus result in an unsafe environment wherever AIs are deployed.[25]

Third, it will be difficult to bring proceedings against AIs or hold them to account. A corporation may employ lawyers or seek outside counsel. AIs do not (yet) have the capacity to argue their case in court, appoint a lawyer to represent their interests, or engage meaningfully with a plaintiff to reach a settlement; furthermore, these capacities do not seem likely soon.

Finally, AIs do not yet have the capacity to suffer, and it is unclear if it is possible to program or develop empathy digitally, such that an AI would meaningfully understand suffering in others. Further, an AI system cannot today interpret it ethical responsibilities on a contextual basis, nor is it intrinsically aware of its own existence.

6. Conditions for AI Personhood

From the above, we distill four conditions for AI Personhood.

Technological We need to be able to delimit the boundaries of a particular AI system, as AIs can integrate and depend on many external systems for their functioning.

Economic If AI Personhood allows for increased innovation and economic growth, we must identify the beneficiaries and how the gains benefit society. Negative externalities and consequences should be understood and accounted for. Instrumental economic reasons must be scrutinized from a diverse range of perspectives.

Legal AI Personhood would be a far-reaching change in society that must not be taken lightly. Arguments from legal efficiency would require evidence that the current law is insufficient. Similarly, a claim that current law retards the development of beneficial AIs must be carefully assessed. Significant justification should be required to enact such a fundamental change to the legal system, and great care should be taken that AI Personhood is not abused by powerful interests.

Moral AIs must begin to function like current legal persons (i.e., individuals, corporations, and nations). In line with Bentham and Singer, we agree that the ability to suffer in some form is essential. On the other hand, we would also argue that it may be considered immoral to create an AI system that can suffer in the first place. In addition, we agree with Kant and Regan that some form of intrinsic worth, such as the ability to reason about one's own existence and moral duties, is critical.

7. Conclusion

In our view, none of these four conditions are met today. The technological trajectory also does not point in the direction that these conditions will be met soon. We can imagine that in the far future AIs may become much more like people, to the point where we are morally compelled to grant them rights and responsibilities. In fact, several movies and

science fiction stories have allowed us to imagine this technological trajectory. However, this trajectory is more difficult to foresee based on the current state of AI research. We therefore do not think that a speculative possibility should affect or legislative decisions and resources today.

8. Acknowledgements

The main ideas of this paper were discussed at a workshop on AI Personhood at the Princeton Center for Information Technology Policy (CITP) on May 11 & 12, 2017. The workshop attendees included the authors, Peter Asaro, Joanna Bryson, Thomas Burri, Vincent Conitzer, Ed Felten, Brett Frischmann, John Havens, Joanny Huey, Konstantinos Karachalios, Ugo Pagallo, Joel Reidenberg, and Yan Shvartzshnaider. Ugo Pagallo contributed significantly to prior drafts of this paper. Mr. Zevenbergen was partially supported by a CITP fellowship, and Dr. Finlayson by ONR contract No. N00014-17-1-2983.

References

1. European Parliament, Resolution (2015/2103(INL)) of 16 Feb 2017.
2. L. B. Solum, *N.C. L. Rev.* **70**, 1231 (1992).
3. S. Chopra and L. White, Artificial agents: Personhood in law and philosophy, in *Proc. 16th Euro. Conf. on Artif. Intell.*, (Valencia, Spain, 2004).
4. J. R. Crawford, *The Creation of States in International Law* (Oxford University Press, Oxford, 2006).
5. S. K. Ripken, *Fordham J. of Corp. & Finan. L.* **15**, 97 (2009).
6. J. C. Gray, *The Nature and Sources of the Law* (Columbia, New York, 1921).
7. J. W. Salmond, *The Theory of the Law* (Steven & Haynes, London, 1902).
8. J. Bentham, *The Collected Works of Jeremy Bentham: An Introduction to the Principles of Morals and Legislation* (Oxford, Oxford, 1789).
9. P. Singer, A utilitarian defense of animal liberation, in *Environmental Ethics: Readings in Theory and Application*, eds. L. P. Pojman and P. Pojman (Cengage Learning, Boston, MA, 1998) pp. 39–46.
10. I. Kant, *The Metaphysics of Morals* (Cambridge, Cambridge, UK, 1996).
11. T. Regan, The case for animal rights, in *Advances in Animal Welfare Science 1986/87*, eds. M. Fox and L. Mickley (Springer, Amsterdam, 1987) p. 179.
12. S. Legg and M. Hutter, A collection of definitions of intelligence, in *Proc. 2007 Conf. on Adv. in Artif. Gen. Intell.*, (IOS, Amsterdam, 2007).
13. J. McCarthy, What is artificial intelligence? (2007), `http://www-formal.stanford.edu/jmc/whatisai.pdf`. Last access Nov. 15, 2017.
14. M. Scherer, *Harv. J. L. & Tech.* **29**, 353 (2016).
15. B. Goertzel and C. Pennachin (eds.), *Artificial General Intelligence* (Springer, Berlin, 2007).
16. T. Hwang, Computational power and the social impact of artificial intelligence (2018), doi:10.2139/ssrn.3147971. Last access 15 Jul 2018.
17. C. E. Karnow, *Berk. Tech. L. J.* **11**, 147 (1996).
18. Council of the European Union, Dir. 85/374/EEC, vol. L 210, 07/08/1985.
19. American Law Institute, *Restatement (3rd) of Torts: Products Liability.* (American Law Institute Publishers, Philadelphia, 1998).
20. L. Sterrett, *Mich. St. Int. L. Rev.* **23**, 885 (2014).
21. M. S. Shapo, *Corn. Int. L. J.* **26**, 279 (1993).
22. V. R. Johnson, *Washington and Lee Law Review* **66**, 523 (2009).
23. American Law Institute, *Restatement (2nd) Torts.* (American Law Institute Publishers, Philadelphia, 1977).
24. American Law Institute, *Restatement (3rd) of Agency.* (American Law Institute Publishers, Philadelphia, 2006).
25. J. J. Bryson, M. E. Diamantis and T. D. Grant, *Artif. Intell. & L.* **25**, 273 (2007).