White Paper

The Ghost in the Machine – Exploring AI Personhood and Policy

ATARC Artificial Intelligence and Data Policy Working Group

February 2023

ATARC

Advanced Technology Academic Research Center

Acknowledgements

On behalf of the Advanced Technology Academic Research Center, I am proud to announce the release of the 2023 White Paper titled **"The Ghost in the Machine – Exploring AI Personhood and Policy"**, authored by the members of the **Artificial Intelligence and Data Policy Working Group**.

I would like to take this opportunity to recognize the following individuals for their contributions:

Sandy Barsky, Oracle

Anthony Boese, Working Group Government Chair

Tom Brown, General Services Administration

Ken Farber, TekSynap

Kyoung-cheol (Casey) Kim, University of Georgia, Department of Public Administration & Policy

Ansgar Koene, EY

Yvette Gibson, General Services Administration

Suman Shukla, Copyright Office, Library of Congress

John Scott, Department of Veterans Affairs

Brian Seborg, University of Maryland Baltimore County Emeritus

Priya Thotakura, Working Group Member

Sincerely,

Tom Suder, Founder

Advanced Technology Academic Research Center (ATARC)

Table of Contents

1. Introduction

Artificial intelligence (AI) systems continue to expand into new domains of reasoning, thought, and expression. The latest public concern over AI technologies focuses on large language models (LLMs) and their ability to generate seemingly original, logical, and well-organized writing on almost any imaginable topic and in a vast number of styles and forms. Concerns also center on generative AI art models, such as Midjourney which has already won awards in visual art competitions[1]. These demonstrations of creativity have stimulated the public conversation around AI now that AI seems to be approaching otherwise "human" pursuits: creativity has historically been considered the exclusive domain of natural humans. These and other advancements in AI and in society's understanding and treatment of AI systems leads to the examination of whether AI systems are to be considered "persons".

While the consensus among AI researchers is clearly that no current system is sentient (and many believe that current technologies are not capable of creating sentience), there is also strong belief among current researchers that AI systems will become as intelligent as humans by 2075[2].

Recognizing that the evolution of public perception (regardless of expert opinion) often drives public policy development, ATARC's AI and Data Policy Working Group has prepared this report, to serve as a discussion of AI personhood, its meanings, and its potential impacts on society. Our purpose in writing this paper is to support policy makers in understanding the issue of AI personhood by providing a clear description of the topic, definitions of some relevant terms, and descriptions of differing perspectives on a few of the supporting sub-topics of AI personhood (e.g., comparable cases of limited personhood, the applicability of current legal frameworks, the relationship of rights to personhood, etc.). We begin with working definitions of 'AI' and 'person'. We then look at the status quo of AIs as persons both domestically and internationally. We next address potential benefits and harms to society that might arise from treating AI systems as legal persons. While we discuss the issue from many angles, this paper presents no specific policy recommendations.

This paper builds upon ATARC's AI and Data Policy Working Group's[3] panel discussion on the topic of AI personhood (27 July 2022), which reviewed currently applicable legal frameworks, potential definitions for the terms, how to effectively discuss AI personhood, and how it might affect society and government. A summarized transcript of the panel and notes about the speakers is included as an appendix to this report.

[1] https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html
[2] https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/
[3] ATARC's Artificial Intelligence (AI) and Data Policy Working Group supports U.S. government leaders in effectively acquiring, developing, and applying emerging AI technologies while maintaining and promoting American values. The group engages emerging AI topics to anticipate, frame, and help focus policy issues as they arise.

2. What is Artificial Intelligence?

Artificial Intelligence (AI) has no universally recognized definition. Nevertheless, a working definition is needed to set the scope of this paper. For one, per Brookings, AI systems are "machines that respond to stimulation consistent with traditional responses from humans, given the human capacity for contemplation, judgment, and intention."[4] Moreover, these responses and the "decisions" that lead to them are traditionally those which would otherwise require a human level of reasoning capacity.[5] The question of intentionality is not resolved with respect to AI systems. The discussion generally revolves around whether a self-conscious agent is necessary for a system to be intentional, or is it sufficient that a selection between options is sufficient to display intentionality.[6] In order to perform decision making and other actions and to learn from stimuli, AI acts based on programmed parameters and a pattern of action defined by trial and error or by creating statistical models of historic decision patterns, which is conceptually similar but structurally different to how natural intelligences learn and operate.

Example Case: The EU

While there is no universal definition for AI, some legal and governance groups have set definitions for their own purposes. This example is from the EU definition of a "smart robot":

- The capacity to acquire autonomy through sensors and/or by exchanging data with its environment and the analysis of those data
- The capacity to learn through experience and interaction
- The form of the robot/AI's physical support
- The capacity to adapt its behavior and actions to the environment
- The absence of biological life

Source: REPORT with recommendations to the Commission on Civil Law Rules on Robotics

27.1.2017 - (2015/2103(INL))

Committee on Legal Affairs
REPORT with recommendations to the Commission on Civil Law Rules on Robotics | A8-0005/2017 | European Parliament (europa.eu)

It is important to note that all extant AI are only able to be developed and operate within narrow, pre-specified cognitive bounds and from finite, often preassembled, universes of information. There is no "General AI", a term used for an AI system that can learn, adapt, understand, and use knowledge to manipulate any context or stimulus the way a human intelligence can. Nevertheless, AIs can and do perform some impressive actions. AI systems are the recognized masters of virtually all mental games, even those in which humans engage in psychological manipulation and deception (e.g., Go, chess, poker, backgammon, Risk, and Diplomacy). In current advances, large language models use inputs such as written text, speech recognition, sentiment analysis, and concept clustering to recognize human languages and generate compelling "new" text.

[4] West, Darrell. "What is Artificial Intelligence" *Brookings.* 2018. https://www.brookings.edu/research/what-is-artificial-intelligence/
[5] Shubhendu, S., & Vijay, J.F. (2013). Applicability of Artificial Intelligence in Different Fields of Life.
[6] West, Darrell and John R. Allen. "How artificial intelligence in changing the world." *Brookings.* 2018 https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/

3. What Is a Person?

It seems likely that the very concept of ascribing personhood to an artificial intelligence seems outlandish to most people. Of course, an AI is not a human being, and there might be philosophical or moral justification for identifying 'persons' as 'humans'. This discussion focuses on legal and policy-based definitions and characterizations of 'person' and 'personhood', because these areas are more explicit and consistent with their use of 'person', their use already extends beyond human persons (e.g., to corporate persons), and because our purpose in this paper is to support policy makers.

More specifically, within jurisprudence there is a distinction between a *natural person* and a *juridical person* – which includes artificial persons. Natural persons are living human persons who are born and who have certain mental and psychological capacities. For example, the authors of this paper are natural persons. Note: not all natural persons are legal persons, especially speaking historically, and, in most societies, humans are not fully vested with total personhood until they reach legal adulthood[7]. In contrast, a juridical person is an entity that is born, made (naturally or artificially), or established and is a person only by merit of the law. Juridical persons have included animals, [8] environmental features,[9] estates, government agencies, and partnerships, but the most common example is the corporate person.

Typically, a juridical person is made so by legal provisions, and these provisions often specify only certain rights and liabilities for that juridical person. Which rights and liabilities apply to the juridical person are driven by the specific legal statute under which these rights are being sought. A well-known example of this is the provision to corporate persons the right that natural persons have to spend on political campaigns without restriction as an extension of their right to free speech, which was established by Citizens United v. the Federal Election Commission (2010). While the ability to engage in unlimited political spending is a significant right, it is still only one right of many automatically afforded to natural persons. However, corporate personhood is far from recent. *Northwestern National Life Insurance Co. v. Riggs* (1906) was the first ruling to explicitly state that corporations are considered legal persons and afforded the same protections of the Fourteenth Amendment as a natural person. However, even

[7] Researchers have indicated that legal personhood "is simply the capacity of a person, system, or legal personhood entity to be recognized by law sufficiently to perform basic legal functions," and that this gives rise to the "capability to own property, enter a contract, file a lawsuit, be named in a lawsuit, serve as a legal principle, and serve as a legal agent" (Shawn Bayern, "The Implications Of Modern Business-Entity Law For The Regulation Of Autonomous Systems," Stanford Technology Law Review, 2015)

[8] For example, in 2022 Ecuador granted rights to wild animals in a ruling from its High Court which stated, "wild species and their individuals have the right not to be hunted, fished, captured, collected, extracted, kept, retained, trafficked, marketed or exchanged." (https://animal.law.harvard.edu/news-article/landmark-ruling/). For another, In 2021 the U.S. District Court for the Southern District of Ohio recognized that animals could be an 'interested person' in regards to 28 U.S.C. § 1782

[9] For example, in September 2022 a lagoon in Spain was accorded legal personhood. (AFP, "Spain Grants Personhood Status To Threatened Lagoon," Barrons, September 21, 2022). Also, in 2017 New Zealand granted legal personhood to a river the Whanganui River. (Nick Perry, "New Zealand River's Personhood Status Offers Hope To Maori," AP News, August 14, 2022). Another example: In 2008 Ecuador granted nature the right to "exist, flourish and evolve." (Articles 10 and 71–74 of the Ecuadorian Constitution) For yet another, though later overturned, in 2019 residents of Toledo, Ohio passed the Lake Erie Bill of Rights law which allowed residents and the City of Toledo to file lawsuits on the behalf of Lake Erie

before that, *Santa Clara County v. Southern Pacific Railroad Company* (1886) extended protection under the Equal Protection Clause to corporations, which itself built on precedent pointing in that direction starting as far back as *Trustees of Dartmouth College v. Woodward* (1819). Over this long period corporate persons have gathered many rights and protections, albeit far fewer than the corpus of rights of natural persons.

Cases like these which have accorded rights to juridical persons must be kept in mind when considering personhood and its implications for an AI. Additionally, we note that most juridical persons involve natural persons as their creators and/or controllers, and so suggest that the relationship of natural persons to their AI creations should also be considered.

4. Are AIs Persons Now?

In the United States AIs are not juridical persons at the time of writing, nor is there any public record of legislation currently under consideration to make them so at the federal level. However, other parts of the world do have some policies and provisions of note.

In 2015 the European Parliament Committee on Legal Affairs established a working group to explore "legal questions related to the development of robotics and artificial intelligence foreseeable in the next 10-15 years." The working group published a study in 2016[10] which was incorporated into a resolution to the Commission on Civil Law Rules on Robotics in 2017.[11] One of the key recommendations (recommendation 59f) is that a new electronic person status should be created by "creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently." However, this was not a universally supported position and in response to the proposed resolution an open letter was signed by 285 European Union robotics and AI experts which stated that "from an ethical and legal perspective, creating a legal personality for a robot is inappropriate whatever the legal status model." Nevertheless, it is an important moment for AI personhood and is highly germane.

In 2017 Saudi Arabia became the first country to grant legal citizenship (but not personhood) to an AI system. The AI is named Sophia and is a humanoid robot with speech recognition and advanced synthetic speech technology that leverages deep learning. In 2017 Sophia was named the United Nations Innovation Champion, in 2018 Sophia visited Armenia and stated that: 'living together, creating solutions that use our collective strength, the best of Artificial Intelligence and the best of humans together to address the world's most pressing issues and achieve the Sustainable Development Goals.'

Sophia and its accomplishments have increased social and legal attention on AI, a trend further amplified by recent discussions surrounding former Google engineer Blake Lemoine's June 2022 statement that LaMDA ("Language Model for Dialog Applications", a Large Language AI Model developed by Google) should be considered conscious, and granted its own rights and personhood.[12]

[10] https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf
[11] https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html
[12] https://www.bbc.com/news/technology-61784011;
https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/

5. How Might Treating AIs as Persons Benefit Society?

Whether one thinks that an AI ought to be considered a person, ought not be a person, or has no opinion at all, AIs becoming persons could have large, unexpected impacts on society. Some possible *benefits* that might be realized include the following:

5.1. Social capacity building

Social capacity is the capacity of social systems to produce services and experiences, including its ability to care for people (for example by providing nutrition and shelter to all, especially the young, the ailing, and the disabled), invest in people (for example its ability to socialize, educate and train individuals), and organize people (for example determining power structures and the allocation of resources, providing safety and security, and social steering).[13] The main limiting factors for social capacity are common: there are only so many hours in the day, so many people to work to those hours, so much food to feed those people, and so on. The introduction of AI systems that can enhance social capacity potentially overcomes these limitations with an effective, capable, and complementary workforce. AI personhood would support integrating this effort (on the part of the AI systems) within society through a transparent, modifiable, and democratically determined legal framework. If, in addition to personhood, AI systems were granted self-determination, it may be more likely that AI systems would penetrate more broadly across geographical and economic groups. This may lead to more equitable social capacity building overall.

5.2. More equitable wealth distribution

As mentioned above, the integration of AI to an environment increases the capacity of that environment. This is as true for personal and corporate capacity as it is for social capacity. Under current laws and policies, AI, like any other property, can be accumulated by those with the most capital and used by them nearly however they please. This increase of capacity is expected to generate further wealth accumulation, empowering further AI acquisition, and onward in an accelerating cycle. Meanwhile, all those who were unable to invest as much in AI in the early days of its growth will be left farther and farther behind. This invariably leads to increased economic inequality.

In contrast, if those AIs are recognized as persons and thereby gain the rights to and benefits of their production, then the value they generate would not simply cycle back to their owners, but would instead disperse much more broadly into the economy. Moreover, should AIs become generally wealthier than humans due to their higher capacity for production, this could be a societal benefit. Successful General AI (AI that is capable of human-like, generalized understanding and decision making, as opposed to the "Narrow AI" of today that is able to address only specific tasks) will likely be free from many human irrationalities and biases, and, given their potentially unlimited "lifespans" and far more experienced time (i.e. awake time) during that lifespan, will have dramatically higher potential for generating wealth. If General AI systems were to become the custodians of vast wealth, society may benefit from their ability to apply advanced intelligence to achieve their goals. The risk is that we have

[13] https://www.hq.nasa.gov/iwgsdi/Social_Capacities.html

no current way of anticipating what those goals might be, or mitigating the risks of General AI seeking to achieve those goals with autonomy and the legal rights afforded them due to their personhood.

5.3. Avoid harming sentient entities

The current progress in AI development has not yet established the existential nature of the AI systems themselves. We have not agreed on a definition of intelligence, AI, or consciousness, although they are all closely interrelated and interdependent. Although few companies or labs are directly pursuing the creation of General AI, all AI development can be seen as contributing to the overall progress toward such systems. We can acknowledge the following points:

- even if we believe it unlikely, it is possible that General AI systems will be intrinsically conscious, just as humans are
- our society has generally granted rights to things in proportion to their perceived levels of consciousness – we see humans as having the most inherent rights, pets and livestock have some rights, and inanimate objects are granted rights only in exceptional cases
- we cannot guarantee that the emergence of General AI systems would be immediately known to society at large, the government, or academia

If the legal framework, policies, and laws are not in place to support the rights of these conscious entities, those rights would effectively not exist. Having the legal mechanisms in place to support AI personhood would mitigate against the abuse and harm these systems may experience, were they to arise.

5.4. Steps in the right direction

Above we mentioned personhood for corporations, animals, and natural features. While very different things, the arguments and precedent that help generate rights for one can be used to support generating rights for others. Movement toward granting AI personhood could lead to increased legal respect and protections for more sentient and sapient things. Conversely, not granting AIs person status could lead to fewer legal protections going to inanimate things and social constructs. More directly and especially if done soon, defining and codifying AIs as persons would lay the groundwork so that emergent, sentient AIs arise in a social and legal structure prepared for them and their needs.

White House Blueprint for an AI Bill of Rights

Released in October of 2022, the AI Bill of Rights is a framework addressing the following principles:

- You should be protected from unsafe or ineffective systems.
- You should not face discrimination by algorithms and systems should be used and designed in an equitable way.
- You should be protected from abusive data practices via built-in protections, and you should have agency over how data about you is used.
- You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you.
- You should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems you encounter.

6. How Might Treating AIs as Persons Harm Society?

AIs being treated as persons would introduce a new, ubiquitous class of persons into society, literally integrated (in many cases) in the objects we use and services we access. Possible *harms* that might be realized include the following:

6.1. AI may excel beyond natural persons and become an existential threat

As discussed above, AI systems will be designed for dramatically increased productivity compared to humans. This productivity will, in part, arise from their access to vastly more information than humans, in forms and at speeds with which humans will not be able to compete. If granted property and wealth rights, AI systems would naturally accumulate wealth at a rate far greater than humans or human-owned corporations. This combination of wealth, information, and productivity would grant AI persons tremendous capabilities to fulfill their goals. The issue of whether we will be able to predict, control, or even know the motivations of AI systems has been discussed at length[14], and there is little certainty that their values and priorities would align with our own. Thus, allowing AIs to amass wealth as they become ever more intelligent and capable may result in an existential threat to humanity.

6.2. Natural persons may be displaced

While the idea of racial displacement (i.e., the forced relocation of a group based on perceived racial membership or characteristics) is fundamentally unjust, when confronted with competition from AI systems (which require fewer resources, are tireless, and are likely less emotive) persons, it is not unreasonable to imagine that natural human persons may be displaced from social influence, economic position, and institutional or political power. The degree of social instability that such a dynamic would generate is difficult to anticipate but may certainly be significant even if not to the level of existential threat discussed above.

6.3. Accountability and responsibility may be decreased

AIs that are persons employed or otherwise acting on behalf of a corporation could be another layer of social and legal protection for corporations (in addition to their current advantage in resources and litigious capacity), some of which might intentionally exploit this insulation. Indeed, any bad actors and even merely negligent parties could put into motion a harmful act but do so via one or more layers of AI actors thus insuring any civil or criminal fallout strikes those AIs instead of, or at least before, the corporations or humans are held accountable. Given the relative ease with which AI could be generated, exploited, eliminated, and then disavowed, this would significantly interfere with holding bad human and organizational actors accountable for legal or ethical violations.

Of course, product liability, including strict liability, may still be accorded to the manufacturer of the AI for errors or omissions in the AI that leads to harm. However, in the scenario where an AI or smart machine can make autonomous decisions, current legal precedent may not be sufficient to properly identify the party responsible for providing compensation based upon ultimate causal responsibility,

[14] https://nickbostrom.com/superintelligence

which intuitively seems to lie with the creator or programmer that created the AI with the latent possibility to harm.

6.4. False equivalence between 'humanlike' and 'intelligent'

The current conversation about AI persons often discusses how similar (or dissimilar) the AIs are to humans in their intake, processing, and expression of information. While these topics were also common in early discussions of animal personhood they are no longer, and certainly the apparent humanness of corporations or rivers was never considered in granting them limited personhood. If this trend continues, and especially if some direct or implied equivalence is codified in law, then other potential candidates for limited personhood, such as non-human animals, could be pushed farther from inclusion in the legal category of 'person'. [15] While not every reader will agree that non-human animals should be persons, we can likely agree that such a decision should not be precluded based on personhood decision made regarding AIs as persons.

6.5. Under-specification of the category 'person'

Creating categories allows us to pick out the right sorts of things (for discussion, for legal treatment, for development, etc.) without being too inclusive. Being too inclusive makes the category ambiguous and uninformative, while being too exclusive leads to many objects being left out and reduces the benefit of the category. By adding more and more evidently dissimilar things to a category we run the risk of that category becoming *underspecified*, and thus functionally useless because it neither picks out interesting and possibly unique similarities nor provides a useful filter for keeping out sufficiently unlike things. It is possible, given the apparent social and legal usefulness of the current category of 'person', that, by extending the category to include AI systems, we will lose much of the value of the 'person' category. It may be more beneficial and less risky to explore other options for realizing the benefits of AI systems than by considering them to be persons, regardless of their sophistication.

[15] Aside from this core concern, there are other potential issues in correlating 'intelligent' too closely to 'humanlike' especially when 'intelligent' is a matter of mental, psychological, and/or social attributes. The continued tightening of this correlation very quickly starts to preclude the very young, very old, mentally or physically disabled, and neurologically divergent. Such movement presents increasingly high risk of injustice and shortfall in social capacity as discussed above.

7. Closing Remarks

There are aspects of AI personhood that are not explored here in large part because any discussion would be even more speculative than what is already presented. For example, the implications that potentially 'eternal life' would have on any intelligent entity are beyond our reckoning. Similarly, we are unprepared to discuss the implications of an intelligent entity existing without a corporeal form.

Additionally, this paper is also strictly free of recommendations, which precludes discussion about specific strategies and tactics for potential ways forward and their comparative merits. Of course, these areas should be engaged, assayed, and refined by all stakeholders in the evolution and application of AI systems well before policy development.

Finally, it may be that AIs are neither person nor property and instead an entirely separate legal entity.[16] While interesting, exploration of how those entities are to be characterized and integrated into our legal and policy frameworks are also well beyond the expertise of this group and scope of this paper. We hope that the areas we have covered are helpful to those in leadership positions throughout government to more clearly address, and more effectively develop policy for, creating and applying AI systems to the benefit of our country.

[16] The EU 2017 resolution on Civil Law Rules on Robotics states "creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently." This statement was not accepted unanimously by the Committee, and motivated a rebuttal opoen letter from over 150 AI experts from industry and academia (http://www.robotics-openletter.eu/)

8. Appendix: Summary of 07/27/2022 ATARC Panel on AI Personhood

8.1. Introduction

Personhood is a complex, emerging issue in AI, and one the Federal government needs to begin assessing. There are few, if any, frameworks for AI personhood the government can use to guide decision-making. This panel discusses AI personhood from the perspectives of current laws, current federal AI policy, philosophy, and academics.

The goal of ATARC's AI and Data Policy Working Group is to help federal policy makers develop and implement policies to promote American values while enabling the development and application of emerging AI technologies. We hope to accomplish this goal by providing a forum for broad stakeholder discussion of AI personhood in order to develop clarity of what AI personhood means, how it might affect society and the government, and what areas of focus should be in developing relevant policy.

ATARC's AI and Data Policy Working Group intends to investigate the policy implications of AI personhood for the U.S. Federal government. The topic of AI personhood, legal, ethical, and otherwise, will become critically important as technology advances, public awareness grows, and the limitations of current policies become apparent.

Pursuant to this, the Working Group held a panel discussion on the topic of AI personhood (27 July 2022). Key points discussed were:

AI Personhood Definition: multiple definitions of "personhood" with respect to AI will likely emerge, being derived initially from current legal frameworks, but these are likely to be incomplete or inaccurate in how they influence the integration of these technologies into society

Social impact: society as a whole seems to view AI ambivalently, as both as a source for entertainment and assistance, as well as a potential threat to security, privacy and safety. Much of this ambivalence seems to be related to an unrealistic and uninformed understanding of the technology itself, which is an issue the tech industry and broader tech community need to address.

Representation: as in many areas of AI technology, application, and impact, the study of AI personhood suffers from a lack of a broad representation of viewpoints, perspectives, and experience. This is another area that the tech industry and community need to focus on, to better align the development and application of the technology with broader societal values.

8.2. Webinar Moderator and Panelists

Ken Farber, Moderator	AI Solution Architect at TekSynap and Industry Chair of ATARC's AI and Data Policy Working Group
Anthony Boese	Presidential Management Fellow and Government Chair of ATARC's AI and Data Policy Working Group. Anthony's background includes a decade in academia focusing in-part on understanding intelligence, autonomy, rights, and personhood as they relate to humans, non-human animals, corporations, and technologies.
Ansgar Koene	Global Ethics and Regulatory Leader at EY helping clients put in place governance frameworks and oversight methodology when introducing AI into an organization
Sandy Barsky	Background in the federal government delivering systems that augment humans automating work using statistical methodology, which has evolved to today's artificial intelligence. While at the U.S. General Services Administration and the Veterans Affairs National Artificial Intelligence Institute, Sandy contributed in the creation of the Artificial Intelligence Primer and the Ethical Application of Artificial Intelligence Framework
David Gunkel	Professor at Northern Illinois University researching AI personhood since 2006. Author of three books on the subject: The Machine Question, Robot Rights, and soon to be published Person Thing Robot: a Moral and Legal Ontology for the 21st Century and Beyond

8.4. Q: What do we mean by AI Personhood?

[Boese] I started my research on animals and people trying to figure out what we meant by person, and then [started researching] corporate personhood. We expect that it goes back to the Southern Pacific Railroad in 1898 or so, but some of the major events in the recent past led me to think about corporate persons and technological persons. The change to the email corporations and technology was hard for me, because I was a big fan of suffering-based accounts of what a person was, [that a] person suffers. Corporate suffering seems a little strange to me, clearly there's a line between a moral and legal person. We all respect that [these are] two very separate conversations.

The question of 'what is the legal person' is of course much clearer. Laws are legal constructs, and legal persons are legal constructs with social respect. Ansgar is able to speak more on whether AI should be truly a legal person. But when it comes to a moral or social person, the question becomes a little fuzzier. It's probably hard to suggest that a machine suffers, although maybe. I do try to push back against my opinion about what we understand suffering and communication to be.

Perhaps some of the thinking around planning and continuity that started to come into personhood discussions in the mid-2000s can be relevant to an AI. Even if an AI doesn't have a soul in a way that some people might want, or even if we don't think that an AI can suffer, it does seem like an AI might be the sort of thing that can have a certain sense of a self - a thing that other things are not. The continuity between this self and a future self, where it is the thing that others are not. [An AI sense of self] has some form of plan, even if it doesn't have any emotional investment in it, per se.

For a plan of continuity between the current self and the future self, that identity persists through time and that continuity of planning for time point B, and then time point B becoming realized and being able to reflect back on time point A, is a sort of understanding of a person planning and thinking that I think and AI can probably get to.

Then the question is, why do we care? [If AI is] some sort of moral person in the sense that time continuous, does it gain responsibility because of this fact? Can we morally affront it because of this fact? That's a separate and much more complicated question.

[Koene] In my previous career in academia, I worked on humanoid robotics trying to understand whether our models of how the brain works actually functionally operationalize well. We tried reinforcement learning, [similar to] the type of cumulative learning that you see in children, as a way of enabling robots to gain more capabilities. However, if we look at the state of this type of technology, whether embodied robotics or in software non-embodied kinds of AI systems, we are very far from truly autonomous operating systems.

The idea of a sudden leap to surpass the current limits and accelerate towards super intelligence does not seem realistic to me. The focus of conversations around AI personhood at this stage is around legal personhood, similar to what we see in personhood of organizations. In my research

coming up to this talk, I found a reference to Pope Innocent the Fourth in the 13th century giving personhood to a monastery, so it's not a completely new idea.

It's really a question about where liability lies, and where intellectual property rights, property rights, or the ability to maintain wealth come into play? One of the things that drives conversation around AI personhood at the EU are the challenges for legislative frameworks. The question is whether we need a new legal framework, or whether AI is actually similar to other kinds of tools, like a car, where liability simply lies with the person who is using or who built the system.

The reason why that is a question when it comes to AI is because of things like machine learning. There are actual properties of the system developed over time. You cannot point to an original developer and say all of the system's behaviors are as the developer defined. Those behaviors changed as it acquired new input while it was running. Does that mean the person who was using the system and fed it new kinds of data is the one who should be liable?

One proposal was to shortcut the whole problem of identifying the liable party by giving legal personhood to the AI, thereby giving it the liability. The reason why I don't think that is particularly useful is because the AI does not hold value. It wouldn't be accumulating money for itself, and you wouldn't have a bank account for the AI. It doesn't have the psychological side of personhood. It doesn't choose which tasks to do. You are the one deploying it to do something. So assigning personhood to the AI doesn't really address the question about liability.

The other side is not just about liability, but also about things like intellectual property rights. If an AI system is used to create something new, does the AI start to own the intellectual property around that? Is that sensible? The point of an intellectual property is to make it a valuable area of work for somebody to be engaged in. Someone spent a lot of time and money to create something, which is why we have intellectual property to recoup costs. But if the AI is holding that intellectual property, then what does it gain?

From my perspective, assigning personhood to an AI doesn't really address the challenges that we are actually facing when it comes to AI. To really address the legal issues, you need to go to the next step and cut out AI personhood.

[Barsky] It's very interesting talking about [Pope] Innocence the Fourth, because corporations are granted rights based on moral and religious choices also. We're making this assumption that a corporation can have a religious affiliation and a moral affiliation, [which] goes back to the early church in its use of corporate structures in order to protect property, for instance the monastery.

There was an exercise we did while I was in government when we moved to digital [and] away from print. We found that all the existing laws, regulations and policies could be extrapolated to our new and emerging digital environment. At that time, we didn't see a need for anything new.

People sometimes think differently, and we're going to have to factor in these concepts, not just the moral and religious, but also how our human minds work. What we consider to be feelings, the capabilities to plan, how we see color. If you're a fan of Van Gogh, you might find it very

interesting to see how he saw the world, but if he had seen the world differently, would we have that art that we treasure so much today?

[When] defining personhood, [we] should factor in that people's minds work differently, they're wired differently. Not all are the same, but they bring value. So, when we start to extrapolate that out, we're going to have to consider that with machines. Personhood is defined by rights, but perhaps personhood needs to be defined by obligation.

[Gunkel] This is a really crucial question, and I think this is one of these questions where some terminological precision can help us to parse the question [to] make sense of it better. [There are] three things with regards to this precision that we need to bring to our terminology.

The first is a distinction between two different kinds of personhood: a moral or natural personhood, and legal personhood. Moral, natural persons are like you and I, even animals, and usually benchmarked and determined based on natural capacities, like consciousness, sentience, the experience of pain, etc. Human beings are natural persons, animals can be natural persons, but obviously corporations are not natural persons.

Legal personhood is not based on natural capacities. Legal personhood is based on social exigencies, a kind of social honorarium that we extend to make [something] a subject, and not just an object of law. We are legal subjects, animals can also be legal subjects. But this also engages us in extending personhood to corporations, ships, organizations, and perhaps even AI.

Second, where we get into trouble in resolving these questions is that our legal categories are very rigid and limited. We have only two legal categories, by which we can sort persons and things. This goes back to the Roman juris gaius which says, the law is concerned with either persons or things. We're trying to compartmentalize AI into one of these categories. And we're learning that AI seems to resist [both] reification [and] personification. How do we fit [AI] into our legal systems when it doesn't quite fit the categories or the ontological distinctions that we have at our disposal?

Lastly, there's a difference between the can [and the should]. ‘Can’ is an ability, ‘should’ is a moral question, or at least a legal question. So if you ask me can AI be a moral or natural person? I would say the probability is very low, because we'd have to benchmark it in such a way that we would find it sentient or conscious. But can AI be a legal person? Yeah, all we have to do is make a law declaring AI a legal person. The question isn't can [AI] be a person in the legal sphere, but should it be a person.

This is a question for us, not a question about the AI. It's a question about how we decide and want to integrate this technology into our social sphere. What decisions we want to make for our interests, the best organization of our communities and our individual rights and obligations to each other.

[Farber] Thank you, David. I appreciate hearing a number of things [that I] think are useful. We're trying to be more careful about the language [we use regarding the] distinctions between natural and unnatural persons, and then further distinctions between legal and moral persons. Ansgar raises a particular point that's important for our group's interest, that is our legislative framework isn’t efficient… We find ourselves in this situation of calling corporations persons, because they're

doing some things that are kind of like persons, and there's enough interest in letting them do those things.

The AI question, though, is challenging to figure out, because its behavior we can anticipate is going to be more complex, much more human-like and much more important to control. This other question about IP, ownership and liability [brings up other] questions of [whether there is] a risk as we move into and further develop a knowledge-based economy, creating systems that can generate knowledge almost infinitely. Where's that wealth going to accumulate?

8.5. Q: Are we aware of any instances around the world, in any jurisdiction location where AIs have been granted personhood?

[Gunkel] So there's no decision about personhood per se, but there are decisions about legal standing in regards to various claims that move in the direction of personhood. I'll just give two examples. One is the question of IP. Ryan Abbott and the Artificial Inventor Project has been testing the waters of patent law by bringing suit in a number of jurisdictions on behalf of an AI named Davis. What they've done is file a patent that seeks to give the AI the IP over the invention of a food container. This failed in the EU, and failed in the U.S., mainly because, in both the EU and the US patents are only attributable to persons. Because the AI is not a person, it can't be an inventor. However, in other jurisdictions, like in Australia and South Africa, where the stipulation does not exist, the AI has now been named the inventor of this food container system and is now the holder of the IP.

The second place where we see this is in 12 jurisdictions statewide in the U.S. We've now seen laws come into play trying to integrate the personal delivery system robots on our city streets and sidewalks. In the Commonwealth of Virginia and in Pennsylvania, delivery robots have the rights of pedestrians on the street. Now this isn't a decision about the AI or the robot itself being a person, but it is a decision about how we integrate the movement of traffic, and how we decide liability and accountability in the case of an accident if a driver should happen to take out one of these personal delivery robots. According to the law as it's written right now, if you hit a delivery robot, it's tantamount to hitting a human pedestrian.

[Regarding] liability in those situations [where] the AI owns the patent, I believe in Australia and New Zealand, [the AI] is positioned in some way to benefit from that. I don’t know that is has. Ryan Abbot and his firm are named the representative of the interests of the API, so that they can bring suit on behalf of the AI.

[Farber] Ansgar, I was wondering if that is getting to your question about liability? I'm wondering if you've seen or what you might be able to share with us from the discussions that you've been involved with the EU?

[Koene] It's a great example that David raised. If an AI is the patent holder, who is going to do the enforcing? The AI is not going to take the initiative to sue a person. It doesn't have the capacity

to think beyond its task of creating things for the patent…It's the other parties that have been given the ability to enforce it for the AI…[What do we gain] from assigning a patent to the AI?

At the European Parliament, they decided not to pursue AI personhood, but rather to work on different kinds of AI regulations. The EU AI Act is currently in development, for instance, and soon to be published some proposals around AI liability, the Commission is due to come out with that later this year.

I don't really see the purpose of assigning personhood to the AI. It makes more sense to clarify the law and the accountability around what the AI is doing.

[Ken Farber, moderator] That makes sense, and in response to your question about what do we gain by granting AI's copyright or patent, my limited understanding of the United States' situation is [patents] are granted partially in recognition of what we consider natural rights, but also partially in recognition to support commerce, so that a person benefits from that.

8.6. Q: What would be the form or body of an AI?

[Boese] While the other gentlemen were speaking, I was thinking about other avenues and aspects of ethics or political philosophy that can borrow from thinking about rescue ethics. The notion of moral obligation to rescue generally starts when you don't have to give up something of equitable moral value to perform the rescue.

If we're on a literal boat, and the things are drowning, what is it to say my AI is drowning? Unless it's in a robot, and even then, maybe it's impossible to say that your AI is drowning, because it doesn't have a body. Unless you're really bad with your data management, and you happen to have the integrity of your model exist in a single instance in a single drive, even then I don't know that the drive is necessarily the body of the AI.

For this I know movies aren't a great academical resource, and I know Marvel is far from the best of them when it comes to technological faith, but the whole idea of Jarvis continuing to survive even after multiple suits are blown up, because he's flittering through the network from place to place is a convenient way to visualize the fact that an AI just wouldn't have a body.

I don't know that we would necessarily deprive human consciousness or personhood if somehow human consciousness was extracted from a body and put onto a network. I think we would still respect and be interested in whatever that consciousness represents, so I don't know [that having] a body as necessarily prohibitive or problematic to the course of potential personhood, but I do think the answer is, there is no such thing.

[Koene] There is certainly an aspect to embodied AI that tends to lead us to think of it as an actual entity. This is very much in the eye of the beholder. One example is the attention that's been drawn to this so-called android Sophia. I think this is an absolutely terrible case, because it is leading people to project onto it all kinds of capacities that this machine does not have. It is

effectively a puppet, [but] the strings are just electrical wires and electrical impulses as opposed to physical strings.

It's a puppet doing tasks, but because of the way it has been embodied in a puppet that looks like a humanoid female, it has been given Saudi Arabian citizenship. Not personhood but citizenship. I don't know how you do that, separating the two, but it has. It has attracted all kinds of attention that is ethically very concerning. Public and media attention are feeding into the belief that the machine can do more than it really can.

This connects to the language that is used when it comes to AI. The fact that we're even calling that artificial intelligence, when we don't even have a good definition of intelligence to begin with. Never mind figuring out whether this artificial entity has this kind of talent. The term machine learning makes us think that it is acquiring knowledge in the way that humans acquire knowledge, which is absolutely not the case. The language around AI is structured in a way to make us project capabilities that it doesn't have.

It is some of that projecting that's also leading to the sense that we need to be dealing with AI in a way that's different to how we've been dealing with other kinds of tools and frameworks. Machine learning is basically complicated statistics. So why is this type of complicated statistics requiring a different way of dealing with it than the other, otherwise complicated statistics that we've been using?

[Barsky] One of the things that Nevin likes to say is that artificial intelligence today is neither artificial nor intelligent. The systems that I put together at the General Services Administration used sophisticated forms of statistics. To Ansgar's point, right now artificial intelligence is really statistics on steroids. We haven't gotten to where the qubit is replacing our common way of compute power.

What we're going to be concerned with in the federal government is liability. Who is responsible and accountable? It's very hard to thread out the difference between responsibility and accountability, but it all falls into this mix. The Federal Government works with regulations that come out of legislation [that are] broken down to more specific [agency] policies. That's where it's going to impact the government… on the legal side.

[Boese] Something that Ansgar said reminded me of something that's come [while] talking about animals and people. that of things like that. We're going to draw a line between what is human natural intelligence and what is artificial intelligence. We don't have to say, one is better than the other, we can see they're different.

The instance is to justify [saying that] the machine just has strings in a series of electrical impulses [that] make the strings twitch, [which is] a very fancy marionette. I don't know that there's necessarily much difference to our brains that isn't just a series of electrical and chemical strings that are pulling and making stuff happen. Modern science tends to show that our brain starts to do things before we're aware of it, and the notion of concrete decision is usually reinforcement not activation.

We may be 100% correct that AI are just very fancy marionettes, but I don't know if we can then say therefore they're not people... We think about the idea [that] with animals, people think

they are lesser. When we're talking about these reasons why we can't pull up an artificial intelligence, let's make sure we're not accidentally saying things or not accidentally getting ourselves in a double bind, where we allow humans some special privilege, merely because they're human.

[Gunkel] I just wanted to circle back to a question about physical embodiment and boundaries. We asked the question 'what's the physical embodiment of the AI'? We ask the same question of a corporation or an organization. What's the body of the organization? In law, we're already engaging this problem in various ways. The modern corporation is as decentralized in many cases, or, as in material as what we [believe] AI [to be]. Maybe [the analogy] shouldn't be 'humans and animals', but rather the corporation, as the kind of embodiment that we're thinking of.

8.7. Q: Assuming such regulation exists for legal personhood for AI, how would it end?

[Gunkel] I would say that, among the panelists here, I'm the one most in favor of legal personhood for AI. I think it solves a number of problems with the social integration of these newer intelligence social artifacts into our world.

How we do that is a complicated question, but [for me] real challenge is [coming] up with a new legal ontology that doesn't limit us to either person or thing. That comes from the Romans, [and] it's worked for 2000 years. Maybe in the 21st century, we [need to] come up with a more fine-grained, gradient theory of personhood that allows for distinctions we didn't previously have, and that doesn't categorize something as a 'thing' we can use and abuse as we see fit, or another legal subject that is a full recognized legal person. I think that's far too limiting for the kind of challenges we're looking at.

[Farber] That's interesting. I'm a big fan of ontologies in general, but… there seems [to be a] risk of creating a gradient of personhood. Maybe it's related to Tony's point of pull up pull down. You've created this category, and there might be a risk of misclassification. [Maybe that's what] you're referring to Dave, that there's a risk of doing it, but it might be necessary.

[Gunkel] So, if I can make the scenario even more terrifying in some ways, usually, when you have a dichotomy whether false or real thing or person, you figure out the middle ground [of]what's in between. Historically, both in legal philosophy and in moral philosophy, we have had a middle ground of entities that are sometimes persons and sometimes things based on context. Unfortunately they're called slaves.

The Romans slave law recognize slaves as property, but they can also engage in business transactions, on behalf of their master. Therefore, they were persons for the contracting procedures, but they were property in terms of their legal status with regards to who owned them.

A lot of the [AI] legal literature proposals you can only call slavery 2.0, which is a kind of in-between position for the AI. I think that's really a disturbing trend, because not only is that a historically very cumbersome burden [and] tradition to even think about, but the problems this brings up are much greater than the solutions we think we're getting out of it.

[Farber] I do want to address another question about representation. [In the Q&A chat, a webinar participant] notes that she would very much like to see the voice of women of color represented in this conversation. She goes on to note that all the speakers in this panel are white men.

I would very, very much like to engage with women of color, people of all types, people of all perspectives on these questions, because I think they're vital. We're going to make important mistakes, we're going to continue to make mistakes. AI right now has a very significant PR problem, and I think part of the reason is because of a lack of engagement of different perspectives and views.

I have to ask for help to get in touch with folks you think should be involved, [because] I would very much like to have that conversation. This panel was created by happenstance, and it is an issue of access. Which leads us to the conversation we're having now. I would like to work against that issue.

I had a question for the panel about language and this issue around PR. We've got a very strong emerging conversation society wide around AI, and it's not favorable. Would you each share your view on how we might better address the issue we have with public perspective or perception of Ai, and whether you are individually engaged in any way that can help make that a productive conversation?

8.8. Q: Would you each share your view on how we might better address this issue of poor public perspective or perception of AI? Are you individually engaged in any way that can help make that a productive conversation?

[Boese] The popular discourse, anywhere from classrooms to bus stops to Congress, it seems to indicate that there's this [idea] that AI is far more capable and concerning that it could possibly be, because of …the thought that AI is a robotic human.

My office is attempting to work towards [AI education], in the long term. We're trying to build [as public property] an AI powered system that will teach AI, among other things I suppose with better education [and] with better familiarity a situation could fix itself.

[Koene] I think there's a number of strands that lead to the perception around AI being problematic. One of them is that we're too imprecise about what we actually mean when we say AI. One of the biggest criticisms of the EU AI Act is the definition of AI they give is so broad that it captures everything. In a way, that is the definition most people have in their mind when they're talking about AI.

We're not being clear. Where do I start calling something an AI? Does it have to be machine learning in order to be AI or not? We have fights between the technical community and the policy community about how to scope what is considered to be an AI. The general public tends to be more aligned with how the policy community approaches thinking about AI, that is anything where we are doing an automation that touches on some kind of decision making.

As such, there is the sense of moving responsibility to a machine, which means distancing ourselves, the organization, from responsibility. That's something we always need to be suspect of, and one reason to be suspicious of AI. If we put the responsibility on the AI, basically nobody takes responsibility.

There's also the fact that Big Tech is the driver behind AI, and Big Tech is going through a loss of trust. That, amongst other things, is linked to things like the ongoing challenges around data privacy. People feel like their data is being used for purposes other than what they understand. AI is implicated in that, and becomes part of the ecosystem that you can't trust and therefore is problematic.

[Barsky] I'm not really concerned about the larger societal [concerns]. I'm more [concerned] with the ethics. Society is going to evolve, and there [will be] many discussions, movies and books that people will read. But, [what is important are] the ethics of anything we automate, [and as] we start to augment ourselves. Remember, we're augmenting ourselves [with AI], not replacing ourselves.

[One example is using AI to augment data for the radiological community in order] to augment what a radiologist can do, so [perhaps] they can see more and be more accurate [with diagnosing patients]. So when we [begin] to break down the ethical application of anything,...the concern with AI is [to ensure certain groups and people are not excluded from the conversation]. We have to be concerned about the bias in [the] research that is done, and where data comes from.

[Gunkel] AI and AI ethics have a communications problem. We see hype being circulated in the popular media, in particular. A lot of this has to do with journalists not well trained in the technology to be able to talk and write about it to effectively communicate the innovations being developed. We have a real need to bring people up to speed on this technology, and not keep it in the hands of various elite institutions, siloed organizations, and disciplines like AI or computer science.

We're only now beginning to realize the need to teach AI across the curriculum to get people well versed in this technology, [so we can begin to] have these conversations. In democratic governance, you want to hear from everyone, not just the experts. That's something we as educators have to do, but also culturally, we have to take this very seriously.

The field of AI ethics, in particular, has a diversity problem. We should also recognize that, when we say ethics, we are often only talking about Eurocentric thinking about ethics. We're talking about utilitarianism from a European perspective and through a European lens. That's not the only way to do this work, and I think we need to bring in a wider range of perspectives on asking these questions and deliberating on these questions from non Western perspectives that come

either out of indigenous traditions that come from either Confucianism or Buddhism. There's a real push now to study this problem from a wider diversity of perspective.

[Farber] Being cognizant of everyone's time, I'd like to thank you all for your perspectives, and I hope this has been a useful conversation for our friends in government and the broader ATARC audience.