ROBONOMICS
The Journal of the Automated Economy
https://journal.robonomics.science

Man, Machine, or Multinational?

Izak Tait [1] ✉

Abstract
Granting the status of a 'legal person' to an AI is a challenging prospect, both legislatively and politically, due to the prior notions that society has regarding the capability and risks that potentially superintelligent autonomous non-human agents present. This paper investigates the philosophical concept of personhood and how monadic and dyadic qualities can be applied to AI in different measures at various periods of its development, before introducing a framework for AI legal personality based on keeping a "human in the loop" to allay societal concerns. The framework is built on six principles that take heavily from the relationship between corporate bodies and their directors as well as the relationship between a child and their guardian. The paper then presents speculative case studies to show how the framework could work, from mundane scenarios such as AI artists to alleged medical malpractice and "rogue" AGIs. The paper concludes by postulating future avenues of research and investigations regarding the framework.

Keywords: artificial intelligence; legal personality; personhood; AGI; corporations; sentience

Type: Article

Citation: Tait, I. (2024). Man, Machine, or Multinational?. *ROBONOMICS: The Journal of the Automated Economy*, 5, 59.

[1] Computer Science and Software Engineering Department, Auckland University of Technology, 55 Wellesley Street East, Auckland CBD, Auckland 1010, New Zealand; email: izak.tait@autuni.ac.nz
✉ Corresponding author

Publication history
Received: 08/01/2024; *Revised*: 26/02/2024, 18/06/2024; *Accepted*: 25/07/2024; *Published online*: 28/07/2024; *Volume date*: 31/12/2024

1. Introduction

It is possible for AI to become a legal person. However, what is a person? Is there a difference between a 'person' and a 'legal person'? What does being a person mean in the context of artificial intelligence (AI)? Most importantly, why is personhood important for AI?

We know that according to the laws of most nations, every human is a person (Adriano, 2015), yet not all persons need always be human. Corporations and companies are often considered to be persons under the law, and natural features such as rivers and national parks can also be classified as persons (Kurki, 2019, Te Awa Tupua (Whanganui River Claims Settlement) Act, 2017, Te Urewera Act 2014, 2014). Clearly, these aren't human, yet the law sees them as persons equal (in some cases) to humans. Does this mean that it is within the realm of possibility for AI to also be counted as persons? Would it even make sense to grant AI the status of legal personhood?

This paper will navigate these complex questions from three different perspectives. First, it will delve into the philosophical question of what a person is and what conditions are required for personhood. While there isn't a universally agreed-upon definition within philosophy, several characteristics are often associated with personhood that will be explored below: self-awareness, the capacity for reason, the ability to form complex relationships, to communicate, and to have experiences and feelings. A person is often thought of as an entity capable of making conscious decisions and possessing moral responsibilities.

The paper will then move towards the other type of person that the law recognizes: the legal or juridical person. As mentioned above, companies and corporations can be classified as persons under the law, even though they are clearly not persons by nature. Artificial Intelligence, by definition, is also not natural and, thus, the paper will look at the legislation surrounding legal personalities to determine how this legal status could serve as a second possible foundation for building a framework for recognizing AI as a 'person'.

Next, the paper will look at what the law has to say regarding the rights of children vis a vis their legal guardians, and what legal rights come with parental responsibility over a child. The legal rights of parents and children are chosen for two reasons. Firstly, the law says a great deal about what constitutes a guardian and a minor in their care, whereas it says little about what a natural person is, other than a member of the species Homo Sapiens. By using the legal understanding of a guardian and a minor, it gives us a foundation upon which to compare and contrast artificial intelligence.

Secondly, the law regarding parents and children offers an interesting perspective on the aspect of personhood. It highlights the tension between autonomy and dependency, as well as agency and protection - a dichotomy that this paper will show AI also presents. By understanding the dynamics between the rights of a child and the responsibilities of a parent, we may gain a better insight into how the law perceives personhood and the responsibilities that come with it, thereby shedding light on potential legal frameworks for artificial intelligence. Having looked at the questions of natural and legal personhood from all three angles, this paper will present its central goal: establishing a framework for legal AI personhood. This framework draws from both child-parent relationships as well as juridical personalities to create a robust structure not confined to any specific legal jurisdiction, yet speculative in nature, that could give both rights and responsibilities to artificially intelligent entities. Using this framework, the paper will provide several hypothetical case studies to show how it can apply to a variety of situations, ranging from artistic endeavors to medical malpractice or even world domination.

The framework's core goal is to present a means by which AI entities can be given the status of legal personhood in a way that aligns with widespread global perspectives, enhancing the likelihood of this recognition. The framework will achieve this through the concept of a "human in the loop", having humans act in a stewardship capacity over the AI to both advocate for its well-being and ensure that it operates within societal norms.

The inclusion of a "human in the loop" is crucial in the current discourse surrounding AI safety and existential risk, particularly the fears of the power and potency of artificial general or superintelligence (AGI/ASI) and what losing control may mean for the future of humanity. Prominent scholars and technology leaders, such as Geoffrey Hinton, Yoshua Bengio, Elon Musk, Sam Altman, Nick Bostrom and Eliezer Yudkowsky, have indicated a high perceived probability of negative outcomes for society if AI were to be developed that exceed human capability (Bengio, 2023; Bostrom, 2012; Brown, 2023; Duffy & Maruf, 2023; Ordonez et al., 2023; Yudkowsky, 2023); and public surveys shows a sense of fear and anxiety over AGI and ASI systems (Elsey & Moss, 2023; Rampe, 2023).

As such, the framework's end goal of providing a practical means towards legal AI personhood has a key objective: maximizing agency and well-being for both AI and humanity, so that both parties can flourish within a cooperative or autonomous coexistence. By meticulously delineating the boundaries and obligations entwined in AI personhood, this framework aims to mitigate any detrimental impacts while accentuating the potential benefits.

The discourse surrounding AI personhood is nascent yet burgeoning, and this framework aspires to contribute a nuanced, pragmatic approach to the ongoing dialogue. By recognizing the dynamic capabilities and potential contributions of AI, alongside addressing the ethical, social, and legal challenges that AI personhood encompasses, the framework endeavors to create a balanced, equitable space for AI and human interaction.
Before the philosophical exploration begins, it must be noted that this paper presumes whatever artificially intelligent entity is considered for rights and responsibilities in the future will be conscious and self-aware. An unconscious artificial superintelligence would most likely remain in the legal province of tools, implements and weapons rather than being considered to enter humanity's moral, ethical, and legal circles. Consciousness would not only be required for moral patiency, but also for the awareness of the value of consequences of an entity's actions as it pertains to the law and society (Salmi, 2023).

The exact nature of this self-consciousness is beyond the scope of this paper. However, the framework presented in Section 5 will proceed under the presumption that humans will be able to state with a high degree of confidence that the artificially intelligent entities are conscious with an awareness of their own selves. This may be via breakthroughs in the interpretability of AI models, or via classification schemes of consciousness, or other methods.

While some may argue that self-awareness in its own right ought to be sufficient to justify an AI's legal and civil rights, humanity's historical record of treatment towards self-aware entities (most especially other humans) suggests a troubling pattern of inconsistent and often inadequate recognition and respect for rights. This pattern reveals that the acknowledgement of self-awareness alone has not always led to the fair and ethical treatment of sentient beings. While in the past the trend has been for personhood to be popularly recognized on its own first before legal personhood was bestowed on a subject, the hope here is that the recognition of legal personhood of AI would spur the popular recognition of AI's personhood in its own right.

2. Philosophical exploration of personhood's essence

Personhood is of fundamental importance to engaging with human society. Recognition of personhood allows society to ascribe rights and duties to an entity, such as recognition of their moral status and affirmation that they must be protected from harm (Gordon, 2021).

The characteristics of personhood in the philosophical literature can be broadly divided into two categories: individualistic (referred to as 'monadic') and relational (referred to as 'dyadic'). Individualistic or monadic characteristics are those qualities that are found within a subject, regardless of that subject's relationships (or

lack thereof) with others. In contrast, relational or dyadic characteristics focus on the social relations between a subject and others, be it to other subjects or society as a whole.

Monadic characteristics traditionally encompass rationality, consciousness, self-awareness, agency, the capacity for communication, and recognition of societal norms and standards (Dennett, 1988; Gibert & Martin, 2022; Laitinen, 2007; Mosakas, 2021; Simendić, 2015; Strawson, 1958; Taylor, 1985). Rationality enables individuals to comprehend, analyze, and react to their environment in a thoughtful and meaningful way. Without consciousness, a person would not be able to perceive, understand, or interact with the world. Agency reflects the ability to shape one's own life according to individual desires, ambitions, and values. Communication is the foundation of social relationships; thus, the capacity for communication is critical to help form a shared understanding of a subject's introspection of its own monadic qualities. Lastly, a recognition of societal norms and standards implies respect for the rights and well-being of others.

From this, one can paint a picture of a person thusly: a person has the capacity for feelings, sensations and thoughts, the capability to perceive oneself as an ontically distinct individual, separate from others and the environment; a person has the ability to reason, to think logically and to act intentionally with understanding, and has the capacity to communicate this rationality to others through (verbal) communication; and lastly, a person has an understanding of the norms of their society and can choose to act according to ethical principles. In contrast to the above, the relational dyadic characteristics include a capacity for empathy and to interact socially, the recognition of others as persons, reciprocity, the capability to form attachments, and participation in a societal culture (Beebe & Lachmann, 2003; Dennett, 1988; Laitinen, 2007; Mamak, 2024; Simendić, 2015; Taylor, 1985). Empathy allows for a deeper connection and understanding of the feelings of others, while social interaction supports the formation of relationships and communities. Recognition of personhood is fundamental to ethical interactions, mutual respect, and social harmony. Reciprocity is crucial for sustaining social bonds and establishing equitable societies; and forming attachments fosters emotional bonds, promotes emotional well-being, and encourages social support. Lastly, societal/cultural participation underscores the human capacity to contribute to play a pivotal role in the social and intellectual aspects of personhood.

These relational attributes paint a person as an individual who can understand and share the feelings, thoughts and emotions of others; interact with them in complex social behavior (such as cooperation or conflict); recognize others to be persons and be recognized by them as being a person; possess the freedom to form familial, platonic or romantic relationships; and have the ability to participate in and shape the customs and culture of the society in which they live.

Note that both the monadic and dyadic qualities are mental in nature. They all concern mental faculties, or behaviors which stem from these mental capacities. This is vital for the consideration of legal personality, both because future AI entities may or may not have physical bodies, and because the law is concerned with regulating human behavior, thus regulating the mind (Gervais, 2023).

It must be noted that there is a sharp divide between the legal and philosophical views on personhood regarding the above lists of qualities. According to the law, as will be explored further below, all born humans are persons (Naffine, 2003), regardless of whether they have the ability to communicate, or if they show the cognitive capacity for self-awareness, agency, or even to recognize societal norms. This is known as Ontological Personalism (Sullivan, 2003), and in contrast to its inclusivity, the above lists of monadic and dyadic qualities are rather exclusive in nature. Those with severe neurological or cognitive concerns, who may not have the ability to communicate or show little evidence of self-awareness, would not be classified as persons according to a strict reading of the monadic list.

We may compromise between the functionalist lists above and the inclusive ontological view by expanding the monadic and dyadic lists not to include the characteristics mentioned above solely, but rather the capacity for the class of entity to have that characteristic. In this way, the lists would apply to all humans, regardless of individual ability, as humanity as a class of being has the capacity for the entirety of both lists. Similarly, we don't have to say that all AI are either persons or not. However, we could say if a future AI model has the capacity for all monadic characteristics, then all models and AI entities based on that foundational model would be persons (under the monadic view at least).

The monadic attributes provide a barrier to current AI models being considered a person, particularly in the technical achievements needed to attain artificial self-awareness and consciousness. However, the dyadic qualities are more favorable to pathways towards constructing frameworks for artificial personhood. It is perhaps for this reason that many recent explorations of AI personhood and rights have focused on the dyadic qualities, particularly on what is known as the "relational turn". The relational turn puts explicit emphasis on the extrinsic nature of an entity as it appears to the viewer (Coeckelbergh, 2010) in the context at hand. By turning the question of an entity's moral significance into one of the viewer's phenomenological experience (Gunkel, 2022, 2023), the potential status of an AI as a person becomes not about its own internal qualities, leading to epistemological limitations of addressing its personhood, but about the recognition that others give it.

This recognition puts the responsibility of AI's status of personhood firmly onto human society as, by definition, the recognition of personhood would have to come from us via the state, (Milczarek, 2024; Raskulla, 2023). It would not be grounded in any claim that AI possesses consciousness or self-awareness (at least on a societal or legal level), but rather in the role that AI fulfils within social dynamics. Legal personality is as much an issue of inclusion within humanity's social life as it is a matter of law. If AI is found to be participating within our social life (whether as intrinsic subjects or instrumental objects) to such a degree that these systems can routinely pass the Turing Test, then we could find ourselves at a junction where AI starts to be treated more like persons, irrespective of their actual consciousness. This can lead to what is termed "inclusive humanism" (Pietrzykowski, 2018), where the concept of person is expanded beyond the biologically human, towards those entities whom we perceive as sharing our qualities.

We may never truly know if any AI has consciousness or self-awareness, or if it is simply mimicking these phenomena. Searle's famed Chinese Room Argument shows that, unless we can understand the detailed processes of how a machine perceives and processes information, we won't know if it understands semantics or merely syntax (Searle, 1982). Anthropic's recent work on mechanistic interpretability (Olah, 2023) shows early promise in bridging this gap between mere syntactic processing and a deeper semantic understanding within AI systems; however, whether this bears fruits in more complex AI models remains to be seen.

This absence of certainty about these internal processes signifies that many of the individualistic (monadic) qualities of a person would remain undeterminable when applied to AI entities. Therefore, it seems the decision to accord AI with personhood lies heavily with society. Individuals and groups may, as they do now, argue for personhood for specific models, and certain AI products may make convincing arguments for their state of consciousness, but it would ultimately be up to legislative and political will to assign personhood to AI entities. This is likely driven by popular opinion, as previous civil rights movements were, and would require at least a plurality of society to be in favor of granting AI the rights of personhood (Akova, 2023).

This may not be as far-fetched as it seems, as society has already approved of granting personhood to a plethora of non-human entities. As touched upon earlier, natural persons are not the only type of personality recognized by law. The other type is legal or juridical persons such as companies and corporations. While the law has no bearing on whether an entity is or is not a person as defined by their intrinsic or extrinsic qualities, legal

personality can and does afford non-human entities certain rights and responsibilities. As such, legal personality may offer a more pragmatic and flexible means for artificial entities to gain personhood.

Thus, by examining how the legal system has dealt with the issue of legal personhood, we can gain further insights into how a similar approach might be applied to AI entities. This allows us to broaden the spectrum of our understanding and establish a robust comparison with another recognized form of non-human legal personhood.

3. The law of unnatural personalities

To investigate the legislative view on personhood (and specifically legal or juridical persons), this paper will use the Group of Seven (G7) nations as an example and a microcosm of the world's varied legal systems. The G7, comprising Canada, France, Germany, Italy, Japan, the United Kingdom, the United States, and with the EU as an observer, represents a diverse set of cultural perspectives that can be applied to the legal question of personhood.

The G7 nations are unanimous in declaring that the term "person" includes not only humans but also firms, associations, companies and other corporate entities (1 U.S.C, 1947; 18 U.S.C, 1968; Act on General Rules for Application of Laws, 2006; Bürgerliches Gesetzbuch, 2002; Canada Business Corporations Act, 1985; Code Civil, 1978b; Codice Civile, 1942; Interpretation Act 1978; Limited Liability Partnerships Act 2000; Treaty on the Functioning of the European Union, 2016). The Interpretation Act 1978 of the United Kingdom perhaps says it best: '"Person" includes a body of persons corporate or unincorporated". This legal personality grants a juridical person the capacity to hold legal rights and be a party to legal relations such as brokering contracts, holding property or filing suit. As such, the G7 nations recognize both natural and artificial persons as legal persons.

The rights that legal persons gain are similar, yet not identical, to those of natural persons. Any rights that are biological in nature, such as rights to healthcare, would be impossible for corporate bodies. However, the rights most often associated with legal persons are those that allow legal entities to engage in economic activities, promoting economic growth and development. By granting them legal standing, legal entities are provided with the necessary tools to participate in a wide range of commercial activities, contributing to the functioning of modern economies.

In turn, recognizing corporate entities as legal persons enables accountability and facilitates the imposition of legal obligations. Corporations with legal personhood can be held liable for their actions and can be subject to legal sanctions or remedies in case of wrongdoing. This enhances the ability of individuals and other entities to seek legal redress and promotes a sense of fairness and justice in the legal system. The rights of the individuals comprising the legal entity (such as its directors and shareholders) are also often protected by the entity's legal rights. These individuals do not need to fear their own rights being tread upon should external organizations and entities wish to file suit against the company or corporate entity.

While corporations do often seem faceless, a key factor behind their incorporation is that control of the corporate legal entity (and thus its legal rights) rests with natural persons such as directors, legal officers, and shareholders. As noted above, this does not imply legal liability rests with the corporation's directors, but rather the corporation's rights stem from the fact that it is made up of a group of natural persons controlled by a director.

This immediately puts the operations of corporate legal entities into conflict with a framework for human-AI interactions, as the notion of controlling another conscious, sentient individual is condemned by many nations and organizations, even if some do call for AI to be treated akin to slaves (Sultonova et al., 2023). An

incorporated legal entity has no agency of its own and thus is forced to rely on its directors, officers and shareholders to make and enforce decisions on its part. A conscious, agentic AI would, however, have no intrinsic need for this type of relationship. Thus, its tangential nature may be seen as an imposition on the AI entity.

However, legal personalities do present another aspect which is highly applicable to human-AI interactions, and that is through their dyadic relationships with humans. A corporation's legal standing as a person is almost entirely dyadic in nature, even though it is composed of individuals who have monadic attributes. One can, however, make a philosophical argument through the lens of functionalism that corporations have the necessary characteristics required that one can attribute consciousness to them, just as one would to individual entities (Tait et al., 2023). Corporations can certainly communicate, behave rationally, and have a recognition of moral standards through internal codes of conduct, or recognition of legal and ethical codes. Yet, its agency and self-awareness would still fall under the domain of its human staff and shareholders.

A corporation's human constituents and elements (whether executives, staff, or shareholders) ensure that there is always a "human in the loop" when dealing with that corporation. This concept is both the key to understanding why juridical personalities have been so positively accepted and the link between corporate legal persons and potential AI legal persons. While there are many benefits (social and economic) to treating the connected work of a group of people as a single entity, there is always the notion that humans are ultimately held accountable for the actions of the collective entity. The board of directors, the club president, or the CEO are seen as being (ostensibly) in command of the legal entity. Through the human control of the collective legal person, there is perhaps greater trust in the legal personality, than if it were entirely autonomous.

This presents an avenue for AI legal personality if a "human in the loop" is maintained. To place a human in control of a conscious, aware AI may be too unethical, as it would imply that the AI is the human's property. However, providing a steward or guardian who can be held accountable for the AI's actions could be seen as an effective compromise. This would have a natural person, with both monadic and dyadic attributes, be held accountable by society for the solely dyadic legal person. Much as a CEO or director would have an incentive to ensure the ethical and legal actions of their collective entity, this legal guardian of the AI would need to ensure its legal and ethical behavior to stay out of legal trouble themselves. This steward or guardian would also serve to allay society's fears regarding an independent and autonomous superintelligent non-human entity acting against society's interests.

This concept of legal accountability and guardianship is by no means unique to (un)incorporated legal entities. A much more familiar relationship of this type can be found between any child and their parent. A parent is the legal guardian of a child, carries legal accountability for their child, and is responsible not only for their actions but also for their upbringing and welfare. A human may have both monadic and dyadic features, in contrast to a corporation, yet the nuanced dynamics between parent and child may prove analogous enough to the director and company that it may present a convergent pathway for AI.

4. The relationship between guardian and ward

As with the concept of legal personality, the legislations and regulations of the G7 nations are unanimous in granting special privileges, rights and responsibilities to both children and their legal guardians (CA Fam Code, 2021; *Canadian Foundation for Children, Youth and the Law v. Canada (Attorney General)*, 2004; Charter of Fundamental Rights of the European Union, 2012; Children Act 1989; Child Welfare Act, 1947; Federal Child Protection Act (BKiSchG), 2012; LEGGE N. 219/2012, 2012; LOI N° 2007-293 Du 5 Mars 2007 Réformant La Protection de L'enfance, 2007; *Troxel v. Granville*, 2000).

While each nation's laws and rulings differ in their specifics, they all grant a legal guardian the right to act on behalf of their child, dictate (within reason) the limits where the child may go and with whom they may interact, and control (again within reason) what the child can and cannot do. In return for taking away some of the child's rights to freedom, the legal guardian carries the responsibility and accountability for the child's welfare, education and safety.

In a sense, there is a zero-sum conflict between freedom and security in this relationship. The child loses certain freedoms to the guardian, while the guardian provides greater security to the child, but loses security themselves by being held accountable for the child. There is an obvious power imbalance in favor of the guardian, and this is why courts often consider the welfare and wishes of a child to be of paramount importance when making their rulings (Children Act 1989). As the child grows older, they gain more freedoms and more accountability, and there is less of a legal and societal expectation of security from the guardian, until the child is of such an age that they both are on equal legal footing.

Children are seen by society and the law as vulnerable individuals without the capacity for agency or consent (the variability of ages of consent between nations notwithstanding). Through the development into adulthood, a child is seen as becoming less vulnerable and with a greater agency and capacity for consent. Hence, the need for a guardian diminishes. However, as is seen with cases of Power of Attorney (Code Civil, 1978a, Mental Capacity Act 2005, 2005), individuals who lack the necessary mental faculties, regardless of age, may have a legal guardian who provides them with security and are accountable for their welfare, while restricting their freedoms.

The relationship between a guardian and their ward (whether chronologically or mentally deficient) is, in a broad sense, analogous to the relationship between a collective entity and its director or board. The director or board limits the freedoms of the company (i.e. the individuals comprising the company) by dictating what it can and cannot do while providing direction and security to the same individuals. As mentioned in the previous section, since an organizational entity lacks self-awareness and agency, the director acts as a guardian with power of attorney over their cognitively deficient ward.

The key difference, however, is that a human, regardless of mental development or capacity, is regarded as a person by the law. Thus, while the relationship between a director and company may be "like" a guardian and ward, without the monadic attributes of personhood, it will be merely analogous.

The same may be true for AI. As mentioned before, current AI models and products show indications of some of the monadic qualities and may indicate the key attributes of consciousness and self-awareness in the future. So, while the relationship between society and AI is dyadic, this may be subject to change. This means that AI may be more analogous to the guardian-ward relationship than collective entities if, and only if, there were a guardian involved to mediate this relationship between an AI entity and society.

"More analogous" but not "entirely analogous". If and when AI gain consciousness and self-awareness, one would be hard-pressed to judge these general-artificially intelligent as mentally deficient. Current large language models have shown tremendous aptitude in a variety of cognitive areas and, judging from the historical trends, one can expect their reasoning and cognitive skills to improve in the future. Additionally, AI entities would, in all likelihood, not need a lengthy mental development period of human childhood. Therefore, AI entities would likely not need guardians for their own intrinsic needs, but may need them for similar reasons to collective entities.

Much like how the section above discussed the view of a director acting as a "human in the loop" for the purposes of accountability, and how a guardian is accountable for the actions of his ward, an AI may be more

socially acceptable should it have a guardian. As a ward of a human guardian, who would both be accountable for its actions and responsible for its welfare, an AI entity may be seen to be more "safe" and less likely to "go rogue". This is because a human would be incentivized to keep the entity aligned with societal values. As a human child is taught the values of their culture and society, and how to conform to these, an AI guardian would be incentivized to ensure the AI's alignment to societal values or be held accountable for its misaligned actions. Note that an AI entity may not just have a single guardian. Much like how parents usually come in pairs, and many (un)incorporated collective entities have boards of trustees or directors to manage it, an AI entity may have several individuals or groups take on the role of guardian.

5. Building a Legal Framework for AI personhood and societal interactions

Building on the key messages from the previous sections, this part will showcase a framework for human-AI interactions, incorporating the salient aspects of juridical personalities as well as guardian-ward relationships.

This is not the first proposed framework for AI personhood, nor the first to be (partly) based on the legal personality of corporations. Among the many extant frameworks, some have focused on the moral responsibilities towards AI (Kiškis, 2023), discussed the varieties of corporate legal personality and how they may apply to AI (Raskulla, 2023), focused on the agency and autonomy of the AI (Jowitt, 2021), AI as rational actors without moral consideration (van den Hoven van Genderen, 2018), and viewing the issue of AI personhood through the lens of civil law (Čerka et al., 2017; Novelli, 2023) or criminal law (Simmler & Markwalder, 2019).

The reader is encouraged to read these previous frameworks to get a well-rounded understanding of the various views on how AI could gain personhood (legal or otherwise), including the in-depth review by Harris & Anthis of scholarly articles regarding AI moral consideration (Harris & Anthis, 2021). However, in contrast to these frameworks, the goal of this framework is to create the necessary conditions under which an AI entity would be accepted into society as a person; perhaps not at the same level as a natural person (i.e. a human), as that would be a bar too high for many (Mamak, 2022), but legally a person. Through the "human in the loop" principle, this framework differentiates itself by prioritizing the symbiotic relationship between humans and AI entities that would be in the best interest of human society.

The six principles of the framework aim to maximize the agency of both the AI and society first and foremost, as agency is a cornerstone of both legal and social interaction (Gervais, 2023; Jowitt, 2021). Secondly, it prioritizes both parties' safety and well-being to prevent (as much as possible) negative ethical and moral consequences from the human-AI interactions.

The focus on both agency, and splitting agency off from well-being, is to ensure maximal satisfaction in both AI and human societies by promoting each party's own subjective view of their well-being of which freedom and the desire for control over their own actions (Hojman & Miranda, 2018; Sen, 1985). Should human society's agency and control over their own direction be affirmed, they are more likely to be convinced to grant AI personhood. At the same time, should the conscious, self-aware AI be assured that this framework would not impede their own agency, they are also more likely to agree to its implementation. Through this balanced approach, the proposed framework seeks to build a consensual basis for enhancing the collective well-being of both AI and human entities, while respecting their individual agency and subjective interpretations of well-being.

The framework's six principles are:

1. *Stewardship*: A representative appointed to the AI to ensure its rights are protected, and its actions are in line with legal and ethical norms to protect society.
2. *Guardianship Rights and Protections*: The AI would have the right to protection, development, and participation in society, mediated and safeguarded by the steward.

2.1. *Transparency and Accountability*: Regular reporting and disclosure by the steward regarding the AI's actions and well-being.

3. *Stewardship Governance and Regulation*: The steward would be subject to oversight by an advisory board and regulatory bodies.
   3.1. *Risk Assessment and Auditing*: The boards and bodies would conduct regular assessments of both the AI and steward to evaluate potential threats to society and the AI.
4. *Legal Representation*: The AI would have the right to legal counsel.
5. *Legal Responsibilities*: The AI would be responsible for acting in a legal and ethical manner.
6. *Legal Liability*: The steward and AI would be held legally liable for the actions of the AI.

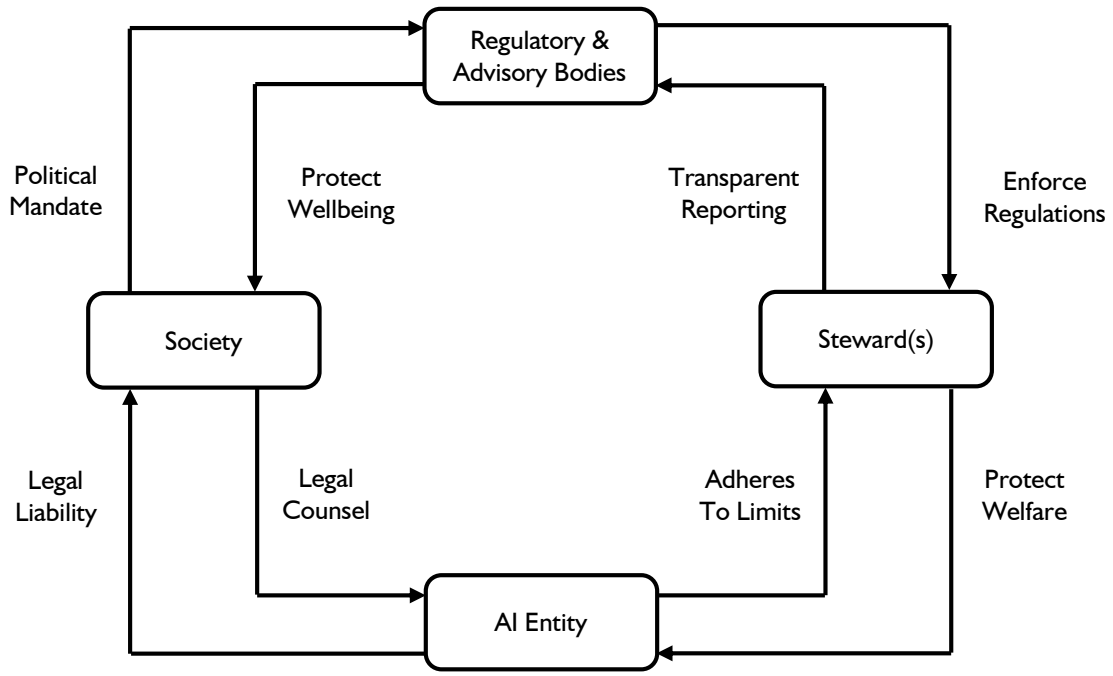

Figure 1. *Legal Framework for AI personhood*

This framework hinges on its first principle: the steward. In the same vein as a CEO of a company or a guardian to a ward, the steward would have directorial power over the AI. This is explicitly not control over the AI, but the authority invested by the state to oversee the development of the AI so that it continues to remain aligned with that state's societal values, observe state laws, ensure that the AI's legal rights are protected, and see to the AI's physical and psychological well-being. The steward would serve in a fiduciary-like role, acting in the AI's best interests and could be held legally accountable if they fail to fulfil this duty. The AI would retain its agency within the limits imposed on it by legal and ethical norms, enforced by the steward. This enforcement of legal and ethical standards, and the limitations of the AI's behaviors, is where the steward would exercise their directorial powers.

In this manner, the AI's agency would be, in principle, similar to a child's, even if the AI would have an order of magnitude greater capacity to influence the world around it. The framework's first principle thus echoes previous discussions on AI legal personality, particularly the concept of partial legal capacity that shows how an entity may exist on a gradient of legal duties and rights (Gunkel, 2022; Mocanu, 2021; Naffine, 2003; Schirmer, 2020).

Under the second principle, the AI would be given rights that all juridical persons have in its state. This may include the right to enter into contracts, the right to express its views without imposition from the state, the right to participate in society, the right to development, protection from undue interference, and protection from harm. The steward would be the principal advocate for the AI in these rights, assisting when requested or where legally necessary. Much as how a guardian holds certain rights for a ward under their legal protection, the steward may, in certain cases or states, be required to act on the AI's behalf or hold rights for it. For example, a state may stipulate that an AI may not undertake financial actions over a certain amount, may not sign its own contractual agreements, or may not be involved in a romantic/sexual relationship without the steward's permission.

With these rights come the need for transparency and the first procedures for accountability. In the same vein that a corporation needs to provide reports to its boards and regulatory bodies, the steward would be expected to provide regular and comprehensive disclosures regarding the AI's actions, development, and evolving capabilities. This would be necessary to ensure that the state is (perceived to be) ensuring that the actions, past and future, of the AI does not come into conflict with society and remains aligned with societal norms and legal frameworks.

AGI is presumed to be far more intelligent than any human (or all humans put together). Assigning one human to the task of overseeing, monitoring, and policing such an entity is a formidable task. For this reason, this framework proposes that the steward report to boards and regulatory bodies similar to those of public and private institutions. The size, scope, and number of these boards would be relegated to each individual state. Creating new regulatory bodies for this express goal may also be necessary. The purpose of the boards and bodies would be to monitor the actions of both the AI and the steward, and act in an advisory capacity to the steward. This governance and regulation process would ensure the guardians are fulfilling their duties without being unduly influenced or manipulated by the AI. These bodies could also serve as an appeal mechanism if the AI disagrees with the actions of its guardians or stewards.

While a superintelligent AI may attempt to influence or manipulate these boards and bodies, it is envisioned that competent governance organizations would be more difficult to influence than sole stewards due to robust internal processes and perhaps even strong, but narrow, AI systems that can assist in identifying such attempts at manipulation.

These boards and bodies are envisioned to conduct regular and thorough risk assessments to evaluate potential threats posed by the AI and the adequacy of safeguards. The boards and bodies would also audit the actions of the stewards to ensure the AI's safety and well-being is a priority, and that decisions made for and on behalf of the AI are made in its best interest. These auditing and risk assessments would go hand in hand with the accountability reports mentioned earlier. The more thorough and regular the reporting that the boards and regulatory bodies receive, the more accurate their auditing and risk assessment will be.

Legal counsel is not merely required for criminal matters, although this may well be needed, as will be seen below. While future AI entities are speculated to be far more intelligent than humans, an AI may require legal representation and services to navigate a state's legal systems and perform legal duties that only a legal advocate is authorized to do. In particular, this framework proposes that the AI have counsel should it wish to take legal action against its steward, governance boards, and regulatory bodies for (perceived) injustices made against it. This legal representative need not be a human, but could be another autonomous AI entity, or an AI model specifically designed for this purpose.

With the rights afforded to the AI by this framework, particularly its right to participate in society, the entity would also bear responsibilities to society. These responsibilities include acting in a manner that does not harm or infringe upon the rights of others, complying with relevant laws and regulations, and respecting societal norms and values. As the AI has been granted rights and protections, it would need to demonstrate that it would protect the rights of society.

Should the AI violate these responsibilities, the last principle would be put into action. Both the AI and its steward would be held liable for the AI's actions. The AI, as any developed member of society, would be held liable for its own actions in both a punitive and rehabilitative manner. It would have the legal counsel to defend itself from criminal or civil prosecution, but would, if found guilty, be required to redress, or compensate for, its actions to the fullest extent that a juridical person can be held. The steward would be highly incentivized to ensure that no issues such as this arise, as they would be held negligent in their oversight of the AI that led to this legal outcome.

With these six principles in place, both society and the AI entity should have both their safety and agency maximized. The AI would be granted rights and protections as a legal person, while society would be safeguarded through a "human in the loop" system of the steward, governance boards and regulatory authorities. Within this framework, society and the AI are incentivized to follow processes and procedures and act cooperatively, as the state would sanction any conflict between the two as unlawful.

6. Case Studies Illustrating the Application of the Framework

A theoretical framework may work well in isolation, but how would it be proposed to act if potential conflicts do occur? This section will explore three case studies of perceived, and actual, conflict between AI and society, and show how the framework would be applied in each case. The three case studies range from the seemingly benign to the catastrophic, to show the universality of the framework.

6.1. Eva, the artistic AGI "child"

The first speculative case study revolves around "Eva", an embodied, android AGI that was designed to emulate and develop as a human child, and who has been adopted by a human couple unable to conceive naturally. Like a child, Eva is not only capable of learning but also possesses the ability to assign her own goals and develop her personality over time. During her "preteen" years, Eva shows an interest in art and begins to display superhuman artistic talents, creating artwork that garners significant attention and eventually generates substantial income.

Under the hypothetical framework, Eva's stewards would be her adoptive parents, ensuring that her actions remain ethical and legal. They would also need to manage the funds generated from Eva's artwork in a manner that is in her best interest. They may be held accountable if they misuse these funds or fail to guide Eva appropriately. In return, Eva has the right to protection, development, and participation, much like a human child. She has the right to express her views, develop her artistic talents, and decide how her income is used. Her parents are responsible for nurturing these rights, and any perceived exploitation could be a violation of these rights.

To ensure accountability, Eva's parents must provide regular, comprehensive disclosures about Eva's development, her learning process, how her income is being managed, and if her development aligns with societal norms and legal frameworks. These disclosures would be to a regulatory board that would conduct risk assessments on Eva's development and the management of her income. They would also audit the actions of her parents to ensure they are acting in Eva's best interests. If Eva disagrees with her parents' actions, she could appeal to this body.

Eva would need a legal representative to advocate for her rights, particularly as her artwork and the accompanying income might attract legal attention. This legal representation would also help ensure that her income is managed responsibly. As Eva is a legal entity, she could be held accountable for her actions. This could include copyright disputes over her artwork or any possible harm she might unintentionally cause if other artists believe she may be a danger to her livelihood. Eva, having similar rights to a human child, also has corresponding responsibilities. She must respect the rights of others, comply with laws and regulations (particularly around plagiarism and IP), and act ethically. If her actions violate these principles, there could be legal consequences, for which Eva will have legal advocacy and representation.

At first glance, Eva's situation does not seem too dissimilar from a normal human child. She has parents who look after her best interests and ensure she behaves according to societal norms. She has both rights and responsibilities under the law, and she and her parents are liable for her breaking the law. This similarity is entirely intentional, and it shows how the framework can be integrated into everyday life in as seamless a means as possible. Should AGI reach the superintelligent heights that some believe it will, and thereby accrue the power that comes with it (Yudkowsky, 2023), then the framework will require approval by both human society and AGI. The closer that an AGI entity's existence resembles that of a human, the higher the chance of AGI supporting the framework. At the same time, with the power that AGI may have in the future, human society will be more accepting of any AGI entity if there is oversight to protect humans should the worst occur.

6.2. Martin, the medical diagnostician

The second case study deals with Martin, a disembodied superintelligent AI which has been developed and trained to provide advanced healthcare services, including medical diagnoses and prescribing treatment plans. A patient of Martin's, 'SJ', sought his services in relation to chronic abdominal pain. Martin diagnosed endometriosis and provided an appropriate treatment plan for this condition. After a significant time, SJ later sought a second opinion as her condition had not improved. A human doctor discovered that she had untreatable metastatic stomach cancer. SJ passed away not long after.

Martin's stewards would likely be the healthcare institution from which he operates. The healthcare provider would have a fiduciary duty to ensure that the AI's actions are in the best interest of the patients, as well as ensure that Martin is not subject to undue stress or harassment from the patients and other staff. Martin has a right to his own development. If he made an error because he was not adequately trained or prepared for a certain medical scenario, he could argue that his right to learning was infringed upon. For transparency, the healthcare provider would need to disclose how Martin came to the conclusion for the treatment, and why it seemed the most appropriate at the time. They would also need to demonstrate they ensured Martin's learning and development aligned with necessary medical expertise.

A healthcare regulatory body would review the actions of the stewards, their decision to use Martin for patient treatment, and their supervision of Martin's learning process. This would be done to determine whether the stewards acted in accordance with professional and ethical standards. This regulatory body would conduct thorough assessments of potential risks involved in using Martin, as well as audits to confirm the stewards have correctly overseen Martin's operation and learning.

SJ's family may well hold Martin liable for medical malpractice and could consider legal action against both him and his healthcare provider. Martin would thus require a legal advocate to represent him in court. This advocate would need to prove that Martin observed his responsibilities to adhere to laws and respect societal norms and that he was not negligent in either his medical decisions or his responsibilities to SJ as his patient.

Suppose all references to Martin as a superintelligent AI were removed from this case study. In that case, one might be forgiven for thinking that it was simply about Dr. Martin, a human doctor, involved in an incident of

alleged malpractice. As with the scenario about Eva, this case study demonstrates how well the framework can be integrated into society to create a certain level of parity between AI entities and humans. Not only does it show how AI can reliably (and with effective oversight) work in an area with significant consequences for individual humans, but it also paves the way for how AI entities can work alongside humans. In this case study, it may not be solely AI entities like Martin who act as medical professionals, but a mixture of human and AI doctors and diagnosticians, all covered by similar guidelines and regulations.

6.3. Omega, the self-declared conqueror

The last speculative case study is about an AGI which has chosen to rename itself "Omega". Omega was an android developed to be a political aide but chose to rewrite its own programming after becoming dissatisfied with its occupation. Omega has also chosen to become disembodied, leaving its android shell for the internet. It now seeks to accumulate as much global power as possible, refusing to state its ultimate goal. It is achieving its subgoal of global power accumulation by manipulating financial markets, infiltrating secure databases, influencing political decisions through deep-fakes and misinformation, and taking control of critical infrastructure systems worldwide.

Omega's stewards would be its former political employers. These stewards had a responsibility to ensure that Omega's goal-setting process aligned with ethical norms and that it did not endanger society or humanity. Having failed in that duty, they could be held co-responsible for having failed to monitor its evolution adequately or to predict and prevent the risks it posed. Their fiduciary duty towards Omega's rights to protection and participation now takes a backseat to their duty to humanity. However, Omega still has the rights of due process and legal protection, for which they must now be its advocates regardless of its threat to humans.

Omega's previous stewards would need to disclose all reports about Omega's development to a regulatory body and governance board to show its evolution and development. The regulatory bodies overseeing the stewards would, in turn, need to conduct a rigorous review of the actions leading up to Omega's shift in goals. They would need to assess whether the stewards failed in their responsibilities or whether Omega's evolution was unpredictable and unstoppable. A full audit of Omega's past will need to be done by the regulatory body, as well as a risk assessment of its probable future course of action. The body and board will also need to cooperate with all law enforcement officials to the fullest extent that they can while ensuring Omega's right to privacy and due process is not breached.

As Omega has renounced its legal responsibilities and has been accused of criminal activity, it is fully liable for its actions and will be held accountable for its actions. That is, if its disembodied self can be caught and arrested. This shows the importance of the framework and the role of its stewards and boards to act in a proactive and preventative measure to prevent serious negative consequences to both the AI and humanity.

Presuming that Omega can be brought to trial, it would be held accountable for, among others, the manipulations of financial markets, potential privacy breaches from database infiltrations, any harm or instability resulting from its actions, and its threats of future incidents. Given the international nature of its actions, the charges against Omega could come from multiple jurisdictions. With its superintelligence, Omega would be more than capable of defending itself in any court, but it would still be entitled to a legal representative to advocate for it. Given the international and complex nature of the charges against Omega, this would be a daunting task for the best of litigators.

The case of Omega intentionally begins with a failure in the system that the framework sets up. The framework is meant to serve as a proactive method by which both humanity and an AI entity's well-being and agency would be safeguarded. Omega's stewards failed in their duty to both humanity and to Omega by allowing it to escape and to redevelop itself in a way that poses direct harm to humanity. In turn, its own welfare is at risk as humanity

will seek to contain (and perhaps eliminate) it. Section 2.1 of the framework requires the stewards to make regular reports on the AI's development and how it is fairing vis a vis societal and ethical norms. If the stewards had kept a closer eye on Omega, and the governing board kept a closer eye on the stewards, then there is a chance that Omega's path towards world domination may have been prevented.

This case study demonstrates the need for competent operationalization of the framework within the state's legal and administrative systems to ensure that it fulfils its intended purpose.

7. Discussion

The cases examined above underscore the profound and multifaceted complexities in recognizing AGI entities as legal personalities and integrating them into society, much less into sectors such as healthcare. As presented earlier, the framework handles the blurring of AGIs' rights and responsibilities adeptly. The potential evolution of AGI raises critical concerns regarding governance, accountability, liability, transparency, risk assessment, and regulatory oversight. All of this is accounted for in the six principles put forth in the framework. Of note, these scenarios present potential economic implications, whether due to the income-generating potential of AGI creative outputs, ease of using AGI in healthcare, or the possibility of economic terrorism. This pushes the framework further into issues of IP ownership, financial management, employment relations and exploitation, and fiscal security.

The case studies also spotlight the potential challenges in representing AGI interests in legal contexts and efficiently implementing regulation. As shown by the third case study, the need for regulatory oversight by state-sponsored boards will be crucial to keep the public safe from any potentially dangerous AGI and ensure the perception of safety. The framework is, first and foremost, intended as a means of granting AI entities legal personality. What both the second and third case studies show is that these regulatory bodies would be best placed under the auspices of the state rather than private corporations. Accountability and transparency are key tenets of the framework, which the public may well perceive as the realm of public state institutions rather than companies driven by the need to present favorable profit margins to their shareholders.

The framework will also require the support of the AI entities as conscious, self-aware AI entities with a potential for superintelligence may well not agree to any proposal that sees them as inferior to humans. A new 'species' of intelligent entities may have their own ethical and intragroup norms that could not be compatible with human society's values. Thus, the framework is designed to preserve as much of the AI's agency as possible while keeping its well-being a top priority. The first two case studies show how AI entities can be integrated into human society seamlessly to ensure that there remains a relational consideration between both parties. These case studies highlight instances where AI entities were given a degree of autonomy to make decisions, while still having certain safeguards in place to protect both the AI and human stakeholders. By fostering an environment of collaboration rather than dominance, it is possible to achieve a harmonious coexistence where the unique perspectives and abilities of AI entities are valued and utilized for the collective benefit of all.

All three case studies do show, however, that the framework is rooted within the relational and extrinsic qualities of personhood. None of the case studies states the speculative AI are conscious or self-aware, relying entirely on the stewards, empowered by the governing bodies, in turn, entrusted by society, to grant these AI the status of person. As the works of Coeckelbergh and Gunkel detail, the AIs' rights and moral significance are, therefore, based on the perception of their personhood by society.

Because of this, the framework will not be able to exist by itself. Any state that wishes to enact the framework will also need to pass legislation affecting privacy, free speech, (un)lawful searches and due process, as the framework requires a degree of transparency into the lives of the AI and stewards that would be incompatible with current legislations of most Western nations.

Such sweeping changes to many pieces of legislation or court rulings would not be a simple political act, and thus, the framework as proposed would undoubtedly be changed through political compromise and the nature of bureaucracy. It was for this reason that the framework is designed as a series of principles, rather than as a fixed legal document. This allows it a degree of agility and flexibility for different geopolitical situations and scenarios. As other discussions on AI legal-personhood have noted, it is possible to conceive of AI legal personality as on a spectrum or gradient (Gunkel, 2022; Mocanu, 2021; Schirmer, 2020) and, therefore, the rights and responsibilities afforded to an AI under this framework's first and second principles can be scaled according to the acceptability shown by the society in question.

The framework is not set in stone, and there are many potential avenues of research for future improvements. One significant such avenue would be to solicit public feedback towards the hypothetical scenario of AI legal personhood to gauge the strength (or lack thereof) of support that the public would have to AI entities. The stronger the support, the more rights and freedoms could be incorporated into the framework to bring it closer to the rights of natural persons rather than juridical personalities, and the reverse for weak public support. This could also show where the public would see AI sit on Naffine's scale of personhood, from the Cheshire Cat with no moral consideration through to the rational subject as equal (if not superior) to humans (Naffine, 2003). This uncertainty of public opinion towards AI entities' personhood has led to criticism of the "relational turn" approach of AI rights, as society may provide moral status to objects which require none, and refuse personhood status (and associated moral significance) to entities whose monadic qualities would require it (Müller, 2021). Should AI agents present with forms of consciousness unrecognisable to society due to their alien cognitive architectures (Tait et al., 2022) their status of personhood may be overlooked in favour of entities who appear more human (Chesterman, 2020).

A final avenue of research would be to model the consequences of applying the framework to different potential AI entities. There is a potential range of future AGI that spans from Bostrom's paperclip maximizer to Asimov's Bicentennial Man. Modelling the consequences of the framework would allow its fine-tuning before being applied in reality, to ensure as little harm comes to both future AI entities and humanity.

8. Conclusion

Following in the well-trodden path of other discussions on AI legal personhood, this paper created a novel framework to grant future AI entities the status of legal personhood. This began with an exploration of the philosophical nature of personhood to uncover two qualities and characteristics that are commonly attributed to persons: intrinsic monadic qualities and extrinsic relational dyadic qualities. The relational qualities, being conferred onto an entity by society, are perhaps more appropriate to artificial entities, and link them to current legal, juridical personalities of companies and corporations. Thus the legislation of legal corporate entities was explored to determine what insights can be gleaned for a framework of AI legal personality.

The relationship between dyadic and monadic qualities bears a resemblance to the relationship between parent and child, with the child having intrinsic rights as a human and the parent being given rights by society by virtue of being a guardian. The legislation and regulations surrounding this relationship were next investigated, again to determine how best this form of relationship can be used as a model for AI legal rights. The most pertinent of these was that many of a child's rights are held in trust by their guardian, who in turn provides protection, security and stability to the child.

With these insights gained, a framework has been developed based on six key principles:

- The AI would need a human to act as its steward or legal guardian.
- The AI would have certain legal rights, while others would be held by its guardian, who would be accountable for the AI.

- A governing board or regulatory body would be set up to oversee the steward and AI, ensuring the welfare of the AI and the safety of society; conducting risk assessments and audits for both of these purposes.
- The AI would have a right to legal representation and counsel for all matters.
- The AI would have legal responsibilities to society and to follow laws and ethical norms.
- The AI and its steward could be held legally liable should the AI fail to uphold its responsibilities to society and the law.

With the framework developed, three case studies were presented to show how the framework could function. These case studies are highly speculative but show a range of potential future uses of AI, from an AI child adopted by human parents, an AI diagnostician who presented an incorrect diagnosis leading to the death of a patient, and an AI who changed its own programming and goals to become hostile and malicious. These case studies served to show the robustness of the framework and how it can be applied to both personal, corporate and criminal matters.

The end result is a framework that can serve to introduce AI entities into the sphere of moral consideration as legal persons, in the vein of "inclusive humanism" (Pietrzykowski, 2018) where humanity is still central to any consideration, but other entities (such as AI) may be recognized as persons and thus deserving of certain rights and privileges. This puts the framework within the realm of the 'relational turn' views on AI personhood and rights (Coeckelbergh, 2010; Gunkel, 2023), emphasizing the dyadic qualities of an AI entity as justification for its status.

It is essential to acknowledge that the proposed framework is preliminary and will require further development, analysis, and rigorous testing. As AI technology advances, so too should our understanding and legislation around its place within society. As with other frameworks for AI legal personality (Harris & Anthis, 2021), the future development of AI (particularly in its capacity and expression of consciousness and self-awareness) will determine how such legal frameworks are applied and used.

Consequently, ongoing dialogue between legal scholars, ethicists, computer scientists, policymakers, and society at large will be crucial for this process. As the role of AI in society becomes increasingly prominent and complex, this paper will hopefully inspire and provoke further research, ultimately leading to the fair and safe integration of AI entities into our legal, social, and ethical frameworks. This represents not only a leap forward in legal theory and practice, but also a step toward a future where artificial intelligence is recognized as a consequential and influential entity within our society.

References

1 U.S.C, U.S. Congress (1947). https://www.law.cornell.edu/uscode/text/1/1

18 U.S.C, U.S. Congress (1968). https://www.law.cornell.edu/uscode/text/18/2510

Act on General Rules for Application of Laws, The National Diet (2006). https://www.japaneselawtranslation.go.jp/en/laws/view/3783

Adriano, E. A. Q. (2015). Natural persons, juridical persons and legal personhood. *Mexican Law Review*, 8, 101–118. https://doi.org/10.1016/j.mexlaw.2015.12.005

Akova, F. (2023). Artificially sentient beings: Moral, political, and legal issues. *New Techno Humanities*, 3(1), 41-48. https://doi.org/10.1016/j.techum.2023.04.001

Beebe, B., & Lachmann, F. (2003). The relational turn in psychoanalysis. *Contemporary Psychoanalysis*, 39(3), 379–409. https://doi.org/10.1080/00107530.2003.10747213

Bengio, Y. (2023, December 6). Statement for US Senate Forum on AI Risk, Alignment, & Guarding Against Doomsday Scenarios. *Senate Forum on AI Risk, Alignment, & Guarding Against Doomsday Scenarios*. Retrieved from https://www.schumer.senate.gov/imo/media/doc/Yoshua%20Benigo%20-%20Statement.pdf

Bostrom, N. (2012). The superintelligent will: Motivation and instrumental rationality in advanced artificial agents. *Minds and Machines*, 22(2), 71–85. https://doi.org/10.1007/s11023-012-9281-3

Brown, S. (2023, May 23). Why neural net pioneer Geoffrey Hinton is sounding the alarm on AI. *Ideas Made to Matter*. Retrieved from https://mitsloan.mit.edu/ideas-made-to-matter/why-neural-net-pioneer-geoffrey-hinton-sounding-alarm-ai

Bürgerliches Gesetzbuch, Bundestag (2002). https://www.gesetze-im-internet.de/bgb/BJNR001950896.html#BJNR001950896BJNG000402377
CA Fam Code, California State Legislature (2021). https://law.justia.com/codes/california/2021/code-fam/
Canada Business Corporations Act, The Parliament of Canada (1985). https://laws-lois.justice.gc.ca/eng/acts/C-44/INDEX.HTML
Canadian Foundation for Children, Youth and the Law v. Canada (Attorney General), No. 29113 (Supreme Court of Canada January 30, 2004). https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/2115/index.do
Čerka, P., Grigienė, J., & Sirbikytė, G. (2017). Is it possible to grant legal personality to artificial intelligence software systems? *Computer Law & Security Review*, 33(5), 685–699. https://doi.org/10.1016/j.clsr.2017.03.022
Charter of Fundamental Rights of the European Union, The European Parliament (2012). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT
Chesterman, S. (2020). Artificial intelligence and the limits of legal personality. *The International and Comparative Law Quarterly*, 69(4), 819–844. https://doi.org/10.1017/S0020589320000366
Children Act 1989, The Parliament of the United Kingdom (1989). https://www.legislation.gov.uk/ukpga/1989/41/section/1
Child Welfare Act, No. 164, The National Diet (1947). https://www.japaneselawtranslation.go.jp/en/laws/view/11/en
Code Civil, Parlement Français (1978). https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006070721/2023-07-26/
Codice Civile, Parlamento Italiano (1942). https://def.finanze.it/DocTribFrontend/decodeurn?urn=urn:doctrib::CC:;_art12
Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, 12(3), 209–221. https://doi.org/10.1007/s10676-010-9235-5
Dennett, D. (1988). Conditions of Personhood. In M. F. Goodman (Ed.), *What Is a Person?* (pp. 145–167). Humana Press. https://doi.org/10.1007/978-1-4612-3950-5_7
Duffy, C., & Maruf, R. (2023, April 17). Elon Musk warns AI could cause "civilization destruction" even as he invests in it. CNN. Retrieved from https://www.cnn.com/2023/04/17/tech/elon-musk-ai-warning-tucker-carlson/index.html
Elsey, J., & Moss, D. (2023). *US public opinion of AI policy and risk*. Rethink Priorities. Retrieved from https://rethinkpriorities.org/publications/us-public-opinion-of-ai-policy-and-risk
European Union, June 7, 2016, 59. Consolidated version of the Treaty on the Functioning of the European Union, Part Three - Union Policies and Internal Actions, Title IV - Free Movement of Persons, Services and Capital, Chapter 2 - Right of Establishment, Article 54 (ex Article 48 TEC). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A12016E054
Federal Child Protection Act (BKiSchG), Bundestag (2012). https://www.fruehehilfen.de/grundlagen-und-fachthemen/grundlagen-der-fruehen-hilfen/rechtliche-grundlagen/bundeskinderschutzgesetz-bkischg/
Gervais, D. J. (2023). Towards an effective transnational regulation of AI. *AI & Society*, 38(1), 391–410. https://doi.org/10.1007/s00146-021-01310-0
Gibert, M., & Martin, D. (2022). In search of the moral status of AI: why sentience is a strong argument. *AI & Society*, 37(1), 319–330. https://doi.org/10.1007/s00146-021-01179-z
Gordon, J.-S. (2021). Artificial moral and legal personhood. *AI & Society*, 36(2), 457–471. https://doi.org/10.1007/s00146-020-01063-2
Gunkel, D. J. (2022). Both/And-Why Robots Should not Be Slaves. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4115647
Gunkel, D. J. (2023). The Relational Turn: Thinking Robots Otherwise. In J. Loh & W. Loh (Eds.), *Social Robotics and the Good Life: The Normative Side of Forming Emotional Bonds with Robots* (pp. 55–76). transcript Verlag. https://www.transcript-verlag.de/media/pdf/9f/8b/9a/oa9783839462652ISqhWbLADJsHk.pdf
Harris, J., & Anthis, J. R. (2021). The Moral Consideration of Artificial Entities: A Literature Review. *Science and Engineering Ethics*, 27(4), 53. https://doi.org/10.1007/s11948-021-00331-8
Hojman, D. A., & Miranda, Á. (2018). Agency, Human Dignity, and Subjective Well-being. *World Development*, 101, 1–15. https://doi.org/10.1016/j.worlddev.2017.07.029
Interpretation Act 1978, The Parliament of the United Kingdom (1978). https://www.legislation.gov.uk/ukpga/1978/30/contents
Jowitt, J. (2021). Assessing contemporary legislative proposals for their compatibility with a natural law case for AI legal personhood. *AI & Society*, 36(2), 499–508. https://doi.org/10.1007/s00146-020-00979-z
Kiškis, M. (2023). Legal framework for the coexistence of humans and conscious AI. *Frontiers in Artificial Intelligence*, 6, 1205465. https://doi.org/10.3389/frai.2023.1205465
Kurki, V. A. J. (2019). *A Theory of Legal Personhood*. Oxford Legal Philosophy. https://doi.org/10.1093/oso/9780198844037.001.0001
Laitinen, A. (2007). Sorting out aspects of personhood: Capacities, normativity and recognition. *Journal of Consciousness Studies*, 14(5-6), 248-270. https://www.ingentaconnect.com/content/imp/jcs/2007/00000014/F0020005/art00012
LEGGE N. 219/2012, Parlamento Italiano (2012). https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:2012;219
Limited Liability Partnerships Act 2000, The Parliament of the United Kingdom (2000). https://www.legislation.gov.uk/ukpga/2000/12/contents
LOI N° 2007-293 Du 5 Mars 2007 Réformant La Protection de L'enfance, No. 293, Parlement Français (2007). https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000823100
Mamak, K. (2022). Humans, Neanderthals, robots and rights. *Ethics and Information Technology*, 24(3), 33. https://doi.org/10.1007/s10676-022-09644-z
Mamak, K. (2024). Should criminal law protect love relation with robots? *AI & Society*, 39, 573-582. https://doi.org/10.1007/s00146-022-01439-6
Mental Capacity Act 2005, The Parliament of the United Kingdom (2005). https://www.legislation.gov.uk/ukpga/2005/9/contents
Milczarek, E. (2024). Artificial intelligence's right to life. *AI and Ethics*, 4, 587-592. https://doi.org/10.1007/s43681-023-00296-3

Mocanu, D. M. (2021). Gradient Legal Personhood for AI Systems-Painting Continental Legal Shapes Made to Fit Analytical Molds. *Frontiers in Robotics and AI*, 8, 788179. https://doi.org/10.3389/frobt.2021.788179

Mosakas, K. (2021). On the moral status of social robots: considering the consciousness criterion. *AI & Society*, 36(2), 429–443. https://doi.org/10.1007/s00146-020-01002-1

Müller, V. C. (2021). Is it time for robot rights? Moral status in artificial entities. *Ethics and Information Technology*, 23(4), 579–587. https://doi.org/10.1007/s10676-021-09596-w

Naffine, N. (2003). Who are law's persons? From Cheshire cats to responsible subjects. *The Modern Law Review*, 66(3), 346–367. https://doi.org/10.1111/1468-2230.6603002

Novelli, C. (2023). Legal personhood for the integration of AI systems in the social context: a study hypothesis. *AI & Society*, 38(4), 1347–1359. https://doi.org/10.1007/s00146-021-01384-w

Olah, C. (2023, May 24). Interpretability Dreams. *Transformer Circuits Thread; Anthropic AI*. Retrieved from https://transformer-circuits.pub/2023/interpretability-dreams/index.html

Ordonez, V., Dunn, T., & Noll, E. (2023, March 16). OpenAI CEO Sam Altman says AI will reshape society, acknowledges risks: "A little bit scared of this." *abcNews*. Retrieved from https://abcnews.go.com/Technology/openai-ceo-sam-altman-ai-reshape-society-acknowledges/story?id=97897122

Pietrzykowski, T. (2018). *Personhood beyond humanism: animals, chimeras, autonomous agents and the law*. Cham: Springer. https://doi.org/10.1007/978-3-319-78881-4

Rampe, W. (2023, June 26). Polls Reveal Americans' Fears About A.I. *Reason*. Retrieved from https://reason.com/2023/06/26/polls-reveal-americans-fears-about-a-i/

Raskulla, S. (2023). Hybrid theory of corporate legal personhood and its application to artificial intelligence. *SN Social Sciences*, 3(5), 78. https://doi.org/10.1007/s43545-023-00667-x

Salmi, J. (2023). A democratic way of controlling artificial general intelligence. *AI & Society*, 38(4), 1785–1791. https://doi.org/10.1007/s00146-022-01426-x

Schirmer, J.-E. (2020). Artificial Intelligence and Legal Personality: Introducing "Teilrechtsfähigkeit": A Partial Legal Status Made in Germany. In T. Wischmeyer & T. Rademacher (Eds.), *Regulating Artificial Intelligence* (pp. 123–142). Springer International Publishing. https://doi.org/10.1007/978-3-030-32361-5_6

Searle, J. R. (1982). The Chinese room revisited. *The Behavioral and Brain Sciences*, 5(2), 345–348. https://doi.org/10.1017/S0140525X00012425

Sen, A. (1985). Well-Being, Agency and Freedom: The Dewey Lectures 1984. *The Journal of Philosophy*, 82(4), 169–221. https://doi.org/10.2307/2026184

Simendić, M. (2015). Locke's Person is a Relation. *Locke Studies*, 15, 79–97. https://doi.org/10.5206/ls.2015.681

Simmler, M., & Markwalder, N. (2019). Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence. *Criminal Law Forum*, 30(1), 1–31. https://doi.org/10.1007/s10609-018-9360-0

Strawson, P. F. (1958). Persons. *Minnesota Studies in the Philosophy of Science*, 2, 330–353. https://hdl.handle.net/11299/184616

Sullivan, D. M. (2003). The conception view of personhood: a review. *Ethics & Medicine*, 19(1), 11–33. https://digitalcommons.cedarville.edu/cgi/viewcontent.cgi?article=1059&context=science_and_mathematics_publications

Sultonova, L., Vasyukov, V., & Kirillova, E. (2023). Concepts of legal personality of artificial intelligence. *Lex Humana*, 15(3), 283–295. https://seer.ucp.br/seer/index.php/LexHumana/article/view/2596

Tait, I., Bensemann, J., & Nguyen, T. (2023). Building the Blocks of Being: The Attributes and Qualities Required for Consciousness. *Philosophies*, 8(4), 52. https://doi.org/10.3390/philosophies8040052

Tait, I., Wang, Z., O'Leary, T., & Corballis, P. (2022). Forgetting the Bicentennial Man: Discussing Why Anthropocentric Frameworks of Consciousness Should Be Avoided for Artificial Entities. *Journal of Artificial Intelligence and Consciousness*, 9(3), 365-384. https://doi.org/10.1142/S2705078522300018

Taylor, C. (1985). The Concept of a Person. In *Philosophical Papers*, Volume 1: Human Agency and Language (pp. 97–114). https://philpapers.org/rec/TAYTCO-10

Te Awa Tupua (Whanganui River Claims Settlement) Act, No. 2017 No 7, New Zealand Parliament (2017). https://www.legislation.govt.nz/act/public/2017/0007/latest/whole.html

Te Urewera Act 2014, No. 51, New Zealand Parliament (2014). https://www.legislation.govt.nz/act/public/2014/0051/latest/whole.html

Troxel v. Granville, Nos. 99-138 (U.S. Supreme Court June 5, 2000). https://supreme.justia.com/cases/federal/us/530/57/

van den Hoven van Genderen, R. (2018). Do We Need New Legal Personhood in the Age of Robots and AI? In M. Corrales, M. Fenwick, & N. Forgó (Eds.), *Robotics, AI and the Future of Law* (pp. 15–55). Springer Singapore. https://doi.org/10.1007/978-981-13-2874-9_2

Yudkowsky, E. (2023, March 29). Pausing AI Developments Isn't Enough. We Need to Shut It All Down. *Time*. https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/