

## ARTIFICIAL INTELLIGENCE SYSTEMS: AN ETHICAL PERSPECTIVE WITH HUMANITARIAN DIMENSION

**Dr. Deepika Pandoi**

Assistant Professor, Institute of Business Management, GLA University, Mathura  
Email – [deepika.pandoi@gla.ac.in](mailto:deepika.pandoi@gla.ac.in)

**Dr. Aruna Dhamija**

Professor, Institute of Business Management, GLA University, Mathura  
Email – [aruna.dhamija@gla.ac.in](mailto:aruna.dhamija@gla.ac.in)

### Abstract

What conditions would have to be met for us to possibly consider artificial intelligence (AI) systems as persons? To address this question, researchers discuss what it would take for AI systems to become metaphysical persons, otherwise referred to as intelligent agents. An intelligent agent is goal-based, acts upon information it perceives from its environment, and is capable of making decisions and choosing a course of action based on its own experiences. Authors make important distinctions in definitions of persons and personhood, such as the distinction between metaphysical personhood, moral personhood, and moral standing. Simultaneously, present Michael Tooley's notion of personhood, focusing on his requirement that an entity have self-consciousness and a concept of itself as a continuing subject of experiences and mental states. In this paper, the researchers tried to distinct between the Person and Artificial Intelligence. Further, trying to understand the connection between Personhood and Moral Standing.

**Keywords** – Person, Artificial Intelligence, Metaphysical Personhood, Moral Personhood and Moral Standing.

### Introduction

It is important that researcher explains the meaning of the term 'person,' and mention other distinctions that are made within this broad term. In ordinary usage, person is sometimes used interchangeably with human beings. For instance, at a crime scene, a police officer might ask a pedestrian standing by, "Did you see the person that did this?" Or a woman might say of her friend, "She is the strongest person I have ever known." This meaning is not what I'm referring to when asking the question of whether AI systems are persons. As Tooley (1983) explains, there are two distinct uses of the term 'person.' On the one hand, you have the usage above, connoting a *biological* meaning that refers to "individuals belonging to our own species, *Homo sapiens*". On the other hand, he notes that we sometimes refer to animals, such as whales, dolphins and primates, and extraterrestrials as persons. When we call nonhuman animals and perhaps robots 'persons,' Tooley (1983) claims that these are "individuals who enjoy something comparable, in relevant respects, to the type of mental life that characterizes normal adult human beings". In this paper, the researchers tried to distinct between the Person and Artificial Intelligence. Further, trying to understand the connection between Personhood and Moral Standing.

## Literature Review

In his book *what is a Person?* Michael Goodman mirrors Tooley by making a distinction between a biological person or human, and a social person. Goodman accepts John Noonan's definition of a "biological human" as one with "a genetic code of a certain kind- genetically identical with *Homo sapiens*" (Goodman, 1988). However, having the genetic code for *Homo sapiens* does not grant an individual entity other things we associate with human beings, such as having rights, duties, privileges, etc. For instance, Goodman gives the example of incarcerated individuals, who are biologically human but have been stripped of some of their rights, especially those who have been sentenced to death. Those entities that we recognize as having rights, duties and privileges are considered "socially human" (Goodman, 1988), which includes membership in the same community as biological humans.

According to Goodman (1988), being socially human is determined solely by behavior and capacities, and not the presumption that an entity is biologically human. Qualifying behavior might include the demonstration of relevant attributes such as rationality, free will, communication, self-motivated action, and other capacities that facilitate the entity's functioning in the community. Therefore, those entities that possess those traits are called persons, even if it "were of a quite distinct species with radically differing genetic coding from us" (Goodman, 1988), or without a genetic code at all.

In jurisprudence, corporations are considered "juridical persons," or artificial persons, which, according to Elvia Adriano, can refer to a collective person, social person, or legal entity that is deemed capable of assuming obligations and holding rights just as "natural persons," which "refers to a human being" with the same capabilities (Adriano, 2015, pp. 365-389).

UgoPagalla notes that quite a few nonhuman entities are granted legal personhood such as corporations, Whanganui River and TeUrewera national park in New Zealand, the Ganges and the Yamuna rivers in India, and Ecuador's entire ecosystem (Pagalla, 2018). Keep in the mind that most of these examples are of legal personhood, which is different from holding legal accountability (Pagalla, 2018). The difference, according to Pagalla, is that legal accountability is given to entities we perceive as agents in contracts and business law, while legal personhood is motivated by "moral status of humans, their intrinsic worth and capability to suffer, their consciousness, and so forth" (Pagalla, 2018). Contemporary thinkers like Lawrence Solum surmise that if we are willing to include AI systems as legal persons, this will force us to completely redefine what it means to be a person: "Given this change in form of life, our concept of a person may change in a way that creates a cleavage between human and person" (Solum, 1992).

## Clarification on Rights

According to the Stanford Encyclopedia Philosophy's entry, "Rights," by Leif Wenar, "Rights are entitlements (not) to perform certain actions, or (not) to be in certain states; or entitlements that others (not) perform certain actions or (not) be in certain states." When discussing the rights that AI systems might possess as potential persons, I am working from a "natural rights" approach.

*Natural rights* are considered a subset of moral rights, which are rights that are “grounded in moral reasons” (Wenar, 2015). In this view, natural rights are a type of moral rights that are ascribed to humans based on their nature and specific features they have “which make respect for certain rights appropriate” (Wenar, 2015), though theorists differ regarding which features are accorded rights.

In this thesis, researchers will take common features that theorists view as important for ascribing rights, which can be plausibly found in potential AI systems such as replicants, making them worthy of consideration for the same moral rights we grant natural persons. There are other rights approaches, as well, including instrumental rights, contractual rights, and legal rights, each of which will approach the question of whether AI systems can claim rights in different ways, which might also bring them to contrasting conclusions.

### **Metaphysical Personhood, Moral Personhood and Moral Standing**

Tooley’s theory of personhood is given in the context of abortion and infanticide, but it still revolves around the basic question of what properties an entity must have in order to be considered a person. It is important to note that his notion of ‘person’ has been embedded in the “moral right to life” (Tooley, 1983, p. 40). In response to Tooley, Tom Beauchamp has argued that Tooley, and other philosophers, make the mistake of assigning moral status to entities with certain cognitive properties without justification. This happens, Beauchamp claims, because thinkers tend to conflate metaphysical personhood with moral personhood (Beauchamp, 1999, p. 314).

According to Beauchamp, theories of *metaphysical personhood* attempt to identify “a set of psychological properties possessed by all and only persons, such as intentionality, self-consciousness, free will, language acquisition, pain reception and emotion” (Beauchamp, 1999, p. 310).

Metaphysical personhood is a normative concept; that is, writers who create lists of cognitive conditions of a person are choosing characteristics that they think distinguish persons from non-persons without regard to an entity’s species, origin, or type. This allows for the consideration of a computer, robot, animal or a divine entity as a person. As Beauchamp explains, the methodology for choosing the cognitive conditions of a person is not something that can be empirically achieved: “Methodologically, the properties of personhood are presumed to be determinable a priori by consulting our shared concept of a person; a theory does not require empirical discovery. The only empirical question is whether an entity in fact satisfies the conceptual conditions” (Beauchamp, 1999, p. 311).

Some of the cognitive conditions that have been suggested by classical and contemporary writers include self-consciousness, capacity to act on reasons, capacity to communicate, and the capacity for free will and rationality. Many of these characteristics are also captured in Daniel Dennett’s rough definition of a metaphysical person as an “intelligent, conscious, feeling agent” (Rorty, 1976). Dennett mentions that rational agency is one of the common “themes” found in theories of metaphysical personhood, especially in the ethical theories of Kant, Aristotle and Rawls (Rorty, 1976).

In contrast, *moral personhood*, as defined by Beauchamp, “indicates individuals who possess properties or capacities such as moral agency and moral motivation” (Beauchamp, 1999, p. 310). Moreover, he assumes that an entity is a moral person if “(1) it is capable of making moral judgments about the rightness and wrongness of actions; and (2) it has motives that can be judged morally” (Beauchamp, 1999, p. 310). He claims having these properties distinguish moral persons from nonmoral persons. It’s also possible for an entity to qualify as a metaphysical person while lacking the properties required for moral personhood or moral standing. According to Beauchamp, to have moral standing means having certain moral rights and protections while a moral agent has moral obligations and responsibilities in addition to moral standing.

Beauchamp (1999) grants that these criteria for moral personhood presuppose certain cognitive capacities, such as the ability to reason and deliberate between right and wrong actions. Thus, these cognitive conditions would have to be defended separately in a general theory of moral personhood, but it is not explicitly stated that metaphysical personhood (e.g., an entity with cognitive capacities such as rationality and will) is sufficient for moral personhood. A theory of moral personhood would still need to argue for the moral value of possessing certain cognitive abilities. Rather, Beauchamp’s focus is on two main theses regarding moral personhood, one of which is most relevant to my discussion. His thesis is that moral personhood, unlike cognitive theories of metaphysical personhood, is sufficient for *moral standing* (i.e., having moral rights or protections) but it is not necessary.

Beauchamp gives a more robust description of moral personhood in the following:

Moral agents are paradigm bearers of moral standing. Any entity qualifying for moral personhood is a member of the moral community and qualifies for its benefits, burdens, protections, and punishments. Moral persons understand moral reciprocity and the communal expectation that they will treat others as moral persons. It is central to the institution of morality itself that moral persons deserve respect and are to be judged as moral agents. Moral persons know that we can condemn their motives and actions, blame them for irresponsible actions, and punish them for immoral behavior. (Beauchamp, 1999, p. 315)

Beauchamp adds that the moral protections that persons are afforded can be extended to other entities who are too “weak and vulnerable who fail to qualify as moral [persons]” (e.g., children under the age of 18, people with mental disabilities and so on) but still qualify as having *moral standing*, which is based on something other than moral personhood (Beauchamp, 1999, p. 315).

Perhaps granting moral standing to such entities by extension says more about us as possessors of moral personhood than the moral status of these other entities who are granted moral standing. For example, corporations are considered persons, and they enjoy the right to buy and sell property, but only because we have agreed to give them that right—a decision not based on consideration of whether or not they satisfy the criteria for metaphysical or moral personhood. Beauchamp might agree that this does happen, and this is likely what he means by saying some entities are granted moral standing *by extension* of the moral community. However, Beauchamp (1999) claims that such decisions are also motivated by entities that have (what he understands as) nonmoral and non-cognitive capacities such as emotion.

Again, Beauchamp (1999) distinguishes moral personhood from moral standing and states that moral standing does not require personhood or possession of cognitive or moral capacities. In

other words, he claims an entity does not need to qualify for either metaphysical or moral personhood to achieve moral standing, so they are not necessary conditions. Moral personhood, but not metaphysical personhood, is sufficient for moral standing, but it is not the only possible sufficient condition.

As stated above, although Beauchamp (1999) claims that moral personhood is sufficient for moral standing, or the ascription of moral rights and protections, it is not required or necessary. In other words, qualifying for moral personhood is enough to have moral standing. However, it is not necessary to satisfy the criteria for moral personhood in order to attain moral rights and protections that come with moral standing. As Beauchamp explains, "Humans too fail to qualify as moral persons if they lack one or more of the conditions of moral personhood. If moral personhood were the sole basis of moral rights (a view I do not hold), then these humans would lack rights--and precisely for the reasons that nonhuman animals would" (Beauchamp, 1999, p. 316).

Beauchamp means that some humans might fail to qualify as moral persons, such as babies and small children, but this does not mean that they lack moral standing, or moral rights and protections. This is why moral personhood cannot be necessary for moral standing.

### **Beauchamp's Critique of Personhood Theories**

Beauchamp (1999) states that contemporary thought regarding personhood has often determined a specific set of psychological, or cognitive, properties that distinguish one as a person, such as intentionality, self-consciousness, free will, emotion and more, in a metaphysical account.

Theorists then infer entities with these nonmoral, cognitive properties qualify as persons. A further conclusion is made about their moral standing, or the rights they have. For instance, Kuhse and Singer (2009) claims that persons have self-consciousness and view themselves as a continuing self. Moreover, he concludes these cognitive abilities give them a right to life.

In other words, it is commonly assumed that if we consider an entity an intelligent, conscious and rational being then they are automatically granted moral standing. However, these two notions are distinct and can be separated when considering whether certain entities are persons. As Beauchamp argues, it is possible for an entity to possess all the relevant features of metaphysical personhood and "lack all the properties requisite for moral personhood" (Beauchamp, 1999, p. 310). He notes it is theoretically possible for metaphysical properties to be possessed by non-human entities that we would not consider moral persons.

As the example insinuates, Beauchamp believes nothing intrinsic to cognitive capacities entails moral concerns: "Capacities of language, rationality, self-consciousness, and the like simply lack an intrinsic connection to moral properties such as moral agency and moral motivation" (Beauchamp, 1999, p. 310). Examples include higher order animals with certain cognitive capacities that we do not consider moral agents, although he does think we might someday make exceptions for nonhuman animals such as great apes and dolphins (Beauchamp, 1999, p. 316). For instance, David DeGrazia notes studies for both these animals showcase cognitive and moral traits, such as emotionally-complex social relationships, self-awareness, individual personalities and rudiments of moral agency (DeGrazia, 1997, p. 302).

In order to grant entities with such cognitive properties moral personhood or moral standing, one

must make two arguments. First, one must state why satisfying these properties is irrelevant for metaphysical personhood. Second, one must argue for the moral implications of achieving metaphysical personhood, such as a 'respect for persons' argument. Ultimately, he claims that no cognitive property or sets of properties "confers moral standing and that metaphysical personhood of this sort is not sufficient for either moral personhood or moral standing" (Beauchamp, 1999, p.314).

Although Beauchamp (1999) desires to separate questions of personhood from morality, it is important to understand why these two subjects often coincide in the literature. As the following will show, many writers are motivated to create personhood theories to address moral issues.

### **The Connection Between Personhood and Moral Concerns**

Thinkers concerned with personhood define personhood in terms of moral concerns, such as the right to life, which might explain why the distinction between metaphysical and moral personhood is sometimes disregarded. "The concept of personhood is generally assumed to have a central place in morality. Nearly every moral theory in the history of (at least Western) philosophy has held that persons possess exclusive or radically superior moral status," explains DeGrazia. "Personhood has often been invoked in efforts to settle moral disputes about the status or proper treatment of individuals whose status was or is in question, such as slaves, mentally impaired humans, fetuses, and nonhuman animals" (DeGrazia, 1997, p. 320).

For example, Mary Anne Warren defines a 'person' as a member of the moral community, the set of beings with full and equal moral rights (e.g., inalienable rights to life, liberty and the pursuit of happiness). The moral community, as defined by Warren, should consist of "all and only people, rather than all and only human beings" (Warren, 1973, p. 54). Any entity has the potential for membership in this moral community if they have the "potential capacity for rational thought" (Warren, 1973, p. 53). Moreover, she suggests that both reasoning and consciousness might be sufficient conditions for personhood, though these two in addition to the ability for self-motivated activity might together be candidates for necessary conditions of personhood. Two main qualifiers for her own list of relevant features of personhood are that 1) an entity does not need to have all these attributes to be a person, and 2) it is not essential that one of these conditions must be necessary.

Warren (1973) believes it is self-evident considering an entity's personhood entails considering it in the moral sense, or in terms of what rights or responsibilities it might have. In doing so, we can see what entities are and are not considered as persons, and what "the decision that a being is or is not a person implies about its moral rights" (Warren, 1973, p. 54). To demonstrate this point, Warren provides the hypothetical example of a space traveler arriving to an unknown planet and meeting novel creatures he has never seen before. To make this thought experience more interesting, imagine that this space traveler is starving from being without food for days and is seriously contemplating whether these creatures will be his next meal.

Warren (1973) suggests that the way he behaves towards these unknown life forms will depend on whether he considers them persons or not: "If he wants to be sure of behaving morally toward these beings, he has to somehow decide whether they are people, and hence have full moral rights, or whether they are the sort of thing which he need not feel guilty about treating as, e.g., a source of food". Perhaps, if the space traveler was making a quick stop and had no intention of

interacting with these new creatures, he would not be prompted to consider whether they are persons.

Regarding AI systems that might be developed in the future, interaction between humans and AI systems will be inevitable as they will be used in all facets of our personal and professional lives. Warren warns that civilization in the next century will have to be prepared to recognize such “highly advanced, self-aware robots or computers,” as well as “intelligent inhabitants of other worlds,” as persons with a potential claim to moral rights (Warren, 1973, p. 57). However, the issue of whether such entities have moral obligations and responsibilities to us is a related but distinct matter. Like Beauchamp’s distinction between moral standing and moral personhood, Warren distinguishes between having moral rights and having moral obligations and responsibilities.

Tooley specifically describes a person as having the “right to life” and claims that “the concept of a person is crucial for the formulation of many basic moral principles” (Kuhse and Singer, 2009, p. 129). According to Tooley, we have strong intuitions about the wrongness of killing a person, and such sentiments stem from the basic moral principle that “destruction of a person is at least prima facie seriously wrong.” The wrongness we associate with killing a person derives from the fact that killing “involves the destruction of a person” (Kuhse and Singer, 2009, p. 130).

Moreover, he states that we feel the same aversion to killing an entity with “comparable or superior mental life to us,” such as a nonhuman animal, which signifies a belief that destroying a person is “prima facie very seriously wrong” (Kuhse and Singer, 2009, p. 130).

### **Michael Tooley on Personhood**

Tooley’s theory of personhood is given in the context of abortion and infanticide, but it still revolves around the basic question of what properties an entity must have in order to be considered a person. As he argues, an entity qualifies as a person “only if it possesses the concept of a self as a continuing subject of experiences and other mental states, and believes that it is itself such a continuing entity” (Tooley, 1983). Tooley bases his notion of personhood on the requirement that a person have self-consciousness.

To support this claim, Tooley starts with the basic question of what is required for an entity to have desires. This is because his notion of ‘person’ includes the right to life, which includes not only continued biological existence but also the right to continue one’s status as “a subject of experiences and other mental states” (Tooley, 1983, p. 44), and the obligations others have to act or refrain from acting in light of an individual’s right.

Moreover, these obligations are “dependent upon the existence of certain desires of the individual to whom the right is ascribed” (Tooley, 1983, p. 44). Tooley asserts that “A has a right to X” is synonymous with the claim that “If A desires X, then others are under a prima facie obligation to refrain from actions that would deprive him of it” (Tooley, 1983, p. 45).

Tooley applies the requirement of consciousness to the ability of an entity to have rights because he interprets desires as not merely behavioristic but as “states necessarily standing in some sort of relationship to states of consciousness” (Tooley, 1983, p. 45). If desire were viewed in terms

of behavior only, then the result would be that ordinary machines could be attributed desires, such as a machine having a desire to be recharged when it searches for an electrical outlet to charge its batteries (Tooley, 1983, p. 45). If we interpret desire as simply behavioristic, Tooley believes we would have to also attribute rights to machines according to his definition. However, he finds this consequence counterintuitive since it is a “conceptual truth that things that lack consciousness, such as ordinary machines, cannot have rights” (Tooley, 1983, p. 45). However, as mentioned earlier, this is not true of legal persons under the law, as corporations are considered artificial persons with certain rights, but Tooley seems to be focused on a natural rights approach.

With this in mind, Tooley revises his previous claim of what it means for an entity to have a right to something as: "A is the sort of thing that is a subject of experiences and other mental states, A is capable of desiring X, and if A does desire X, then others are under a prima facie obligation to refrain from actions that would deprive him of it" (Tooley, 1983, p. 45).

Replacing the term “X” with the right to continue to exist as a subject of experiences and other mental states brings us to Tooley’s definition of what it means to be a person: “A is a subject of experiences and other mental states, A is capable of desiring to continue to exist as a subject of experiences and other mental states, and if A does desire to continue to exist as such an entity, then others are under a prima facie obligation not to prevent him from doing so” (Tooley, 1983, p. 46).

In the final stage, Tooley claims that we simply ask ourselves “What must be the case if something is to be capable of having a desire to continue existing as a subject of experiences and other mental states?” (Tooley, 1983, p. 46). Of course, an entity must have a preexisting concept of itself as a subject of experiences and other mental states in order to desire to continue its existence as such a subject. This is why Tooley sees it as a necessary condition that an entity has the concept of a self as a continuing subject of experiences, and to see itself as that self.

## **Conclusion**

From author’s understanding, Tooley’s notion of personhood, which is presented in his general theory of rights, can be recapitulated as the following:

1. An important, but not sufficient, component of being a person is having the right to life.
2. If one has a right to life, then one should be capable of desiring to continue existing as a subject of experiences and other mental states.
3. To desire such, one must possess a concept of a continuing entity and believe that one is such an entity.
4. If one does not have consciousness of itself as a continuing entity, then one is not capable of having the desire to continue existing as such an entity.
5. Therefore, if one cannot desire to continue as a subject of experiences and other mental states, then one does not have a right to life, an important component of being a person.



Tooley's requirement that a person be a subject of experiences and other mental states implies that persons must have qualitative experiences as part of their self-consciousness. Please note that while we will argue for self-consciousness as a feature of metaphysical personhood, we depart from Tooley in understanding self-consciousness in terms of access consciousness. We argue that this captures the functionality of self-consciousness in intelligent agents and is potentially programmable for AI systems.

## References

- Adriano, E. A. Q. (2015). The natural person, legal entity or juridical person and juridical personality. *Penn St. JL & Int'l Aff.*, 4, 363.
- Beauchamp, T. L. (1999). The failure of theories of personhood. In *Personhood and Health Care* (pp. 59-69). Springer, Dordrecht.
- DeGrazia, D. (1997). Great apes, dolphins, and the concept of personhood. *The Southern Journal of Philosophy*, 35(3), 301-320.
- Goodman, Michael. *What is a Person?* Clifton, Humana Press, 1988.
- Kuhse, H., & Singer, P. (Eds.). (2009). *A companion to bioethics* (p. XIII). Oxford: Wiley-Blackwell.
- Lawrence, S. (1992). Legal personhood for artificial intelligences. *N. Carolina L. Rev.*, 70, 1231.
- Pagallo, U. (2018). Vital, Sophia, and Co.—The quest for the legal personhood of robots. *Information*, 9(9), 230.
- Rorty, A. (Ed.). (1976). *The identities of persons* (Vol. 3). Univ of California Press.
- Tooley, M. (1983). Abortion and infanticide. 120.
- Warren, M. A. (1973). On the moral and legal status of abortion. *The Monist*, 57(1), 43-61.