Second Class Citizens? Questions of Purpose and Membership for AI and Transhumanism
PCTS Meeting November 4, 2017
Braden Molhoek, Ph.D.

Scholars have invested significant time into thinking about the ethics of artificial intelligence (AI). This includes what kinds of tasks or uses for which AI should be adopted, but also how to think about how AI conceives ethics or how to program its ethics. Although I have my own thoughts regarding this, including a naïve belief that casuistry could be a possible ethical method for AI, I have come to the conclusion that in order to think about that question in a meaningful way, a deeper question must be examined first. That question is what makes an AI "good." I am using good not in the sense of good and evil, but good in the sense of purpose. Central to this question is the notion of purpose, *telos*. In exploring what the *telos* of AI could be, I argue that important parallels appear between the relationship between humans and AI and transhumans and humans. The connecting theme is that of the possible creation of second class citizens. The two issues that this theme connections is membership and the effects of unemployment.

AI and automation has the potential to create problems regarding membership in two ways. My analysis will begin with how AI and automation will lead to increased unemployment for people and some of the effects of unemployment. In particular, there are real mental health consequences as well as concerns about self-esteem and identity. Following this, I will examine the relationship between AI, automation, and the concept of "hard" or "unwanted" work and whether machines could be viewed as second class citizens because of the work they do. I then turn to Noreen Herzfeld's work on the Image of God and AI. Using her typologies, I posit a

number of questions regarding the status or needs of AI depending on how intelligence is defined.

Just as AI and automation lead to the possibility of two kinds of second class citizens, I argue that transhumanism has the same capacities. After making the case for how humans and post humans alike could be viewed as second class citizens, I then attempt to bring Herzfeld's typology into the discussion of transhumanism. Finally, using the concerns brought up throughout the paper, I conclude with tentative steps forward in addressing the concerns of membership and the effects of unemployment.

AI, Automation, and Unemployment

With pushes at all levels of government, local, state, and federal to raise the minimum wage or to encourage the development of a living wage, companies are looking to cut both salary costs as well as benefits such as healthcare through any means. AI and the automation of jobs is promising because after the initial cost of building the machine, the ongoing costs should be far less than a human worker. Machines are not subject to the same employment laws that humans are, further reducing costs or increasing efficiency. Touch screen ordering stations can operate nonstop and do not require overtime pay. Automated cars or trailer trucks do not need to stop for sleep or to take care of other biological functions such as eating or elimination of waste.

Although automation promises a decrease in costs and an increase in profits for companies, it raises concerns for the individuals who lose their jobs due to automation. Americans, more than those in other countries, place a great deal of value on work and how it relates to identity. The American dream is built on being able to choose how one works in order to bring prosperity to their family. This narrative also causes a devaluation of people who are not

employed. People feel that such people are "free loaders," "lazy," not contributing to society, and essentially stealing from fellow citizens. Legislation has passed in several states to drug test individuals who receive government assistance, in order to minimize or eliminate fraud, even though these programs cost far more money than any fraud that has been uncovered.

Effects of Unemployment

AI and automation have the potential to contribute to the further stratification of American society, including having a profound effect on those who find their work replaced by machines. A longitudinal survey was done with young Americans from 1979 to 1994, and using the results of these surveys, researchers examined a number of things, including whether there was a relationship between unemployment and symptoms of depression. After taking account of prior symptoms, socioeconomic status, and family background, the length of previous unemployment predicted symptoms of depression significantly, with the value of "P" being less than 0.01.[1] The effect was more profound with men than women, with the author speculating that women might be stigmatized less because the expectation that they were in the work force was lower.[2] And these results were for people between the age of 29 and 37 at the end of the study (1994), with the average period of unemployment being 1.47 years and the range from zero years to thirteen years out of the work force.[3] I do not have data to back this up, but if employment and its loss had that much of an effect on people relatively new to the workforce, I believe the effects

[1] Krysia N. Mossakowski, "The Influence of Past Unemployment Duration on Symptoms of Depression Among Young Women and Men in the United States," American Journal of Public Health 99, no. 10 (October 2009): 1829.
[2] Ibid., 1830.
[3] Ibid., 1828

would be more profound for people who had worked longer, or had the prospect of never returning to the workforce at all.

Symptoms of depression are not the only problems that faces those who are unemployed. A more comprehensive relationship between mental health and unemployment is Jahoda's theory of deprivation. The loss of employment, Jahoda posited, deprives individuals "of the latent functions that employment provides. These functions are time structure, regular shared experiences, information about personal identity, a link with collective purpose, and enforced activity."[4] An Australian study builds upon Jahoda's work to examine how meaningful leisure activity might reduce both deprivation as well as mental distress in times of unemployment. Unlike the study involving symptoms of depression, this study engages adults, both unemployed and those working full time.[5] Other researchers have helped shape this particular study by showing that keeping busy during the day can help mitigate deprivation in young adults, and that active leisure activities increased fulfillment more than passive leisure activities among young adults. Finally, social leisure activities produced different effects than solitary activities.[6]

Waters and Moore hypothesize that those who are unemployed admit to more deprivation, higher levels of depression, and lower self-esteem than those who are unemployed. They also believe that unemployed individuals engage in fewer leisure activities, and that they engage in fewer social activities than solitary activities and that they receive less meaning from these activities than employed individuals, who they believe engage in solitary leisure less than

[4] Lea E. Waters and Kathleen A. Moore, "Reducing latent deprivation during unemployment: The role of meaningful leisure activity," Journal of Occupational and Organizational Psychology 75, no. 1 (March 2002): 15.
[5] Ibid., 16.
[6] Ibid., 16-17.

social leisure activities.[7] Unlike the U.S. study, Waters and Moore found no connection to the effects of unemployment and gender. Their findings suggest that employment status alone predicts sixty-five percent "of the difference in frequency of social and solitary leisure activities, meaning attained through social and solitary leisure activities, perceived deprivation of latent functions and psychological health."[8] It turned out that unemployed persons did engage in more solitary leisure activities and fewer social activities than those who were employed, though loss of income could also be a factor in this case.[9] Unemployment did lead to increased deprivation of Jahoda's latent functions, and this was significant.[10] The most significant of the latent functions was the deprivation of personal identity, which makes sense "if people think of themselves in terms of the job they do."[11] Meaningful leisure activity does reduce the distress of unemployment, and social leisure activities more so than solitary activities[12], which provides a place for constructive work regarding unemployment, identity, and mental health.

AI, Hard Work, and Membership

On the other hand, AI and automation also provide the promise of saving human lives by replacing human workers in situations that are harmful to humans. Jobs that would expose people to harmful chemical, or dangerous working conditions could be done by machines who may not be susceptible to the environmental effects, or can be repaired or replaced more easily. While this is can be positive for people, once the effects of unemployment are overcome, it does raise potential issues for machines, particularly intelligent ones. It has been shown above that

[7] Ibid., 17.
[8] Ibid., 20.
[9] Ibid., 28.
[10] Ibid., 26.
[11] Ibid., 27.
[12] Ibid., 28.

humans make connections between their work and identity, but they also compare their value to others. There are jobs that are seen as undesirable, jobs that are believed to require no particular skills or education. These jobs confer little to no status socially, and these are the kinds of jobs that will mostly likely be done by nonhuman workers. The question becomes, how do we treat these nonhuman workers, especially if they possess human-like intelligence?

Michael Walzer, in his book, *Spheres of Justice: A Defense of Pluralism and Equality*, examines how communities deal with the question of who is given or denied membership. Nations currently allow for people to live within their borders who are not citizens, but Walzer ultimately believes the distinction between immigration and naturalization should be abolished. In other words, he does not see resident aliens as a positive solution, and would rather people who enter a country to live be given the full rights and privileges as citizens, while also agreeing to abide by the same set of rules that citizens do. This is not currently the case, though, and Walzer states that it is possible that "the state controls naturalization strictly, immigration only loosely. Immigrants become resident aliens and, except by special dispensation, nothing more. Why are they admitted? To free the citizens from hard and unpleasant work. Then the state is like a family with live-in servants."[13] He goes on to say that "the rule of citizens over non-citizens, of members over strangers, is probably the most common form of tyranny in human history."[14] Although Walzer is concerned about humans in the context of immigration, naturalization, guest workers, refugees, etc., I believe his concerns can extend to AI workers as well.

[13] Michael Walzer, Spheres of Justice: a Defense of Pluralism and Equality, Reprint ed. (New York: Basic Books, 1984), 52.
[14] Ibid, 62.

As I said earlier in this section, nonhuman workers will likely have to do the work that humans do not want to do. If this automation is done by machines or computers that are not intelligent, the concerns are far fewer. The comparisons made in that case would be with existing technology that makes work easier for humans, or at the most controversial, the use of animal labor. If, on the other hand, AI performs these tasks, I argue that Walzer's concerns about membership are relevant. If AI is comparable to human intelligence, then I think humans need to think about what place such workers should have in society. If their intelligence is like humans, will they also make similar associations between work and identity that humans do? I simply raise this question now, saving any attempt at an answer for the final section of the paper.

***Imago Dei* and Intelligence: Herzfeld's Typologies**

Noreen Herzfeld, in her book, *In Our Image: Artificial Intelligence and the Human Spirit*, argues that the quest for AI says a great deal about our own nature. She also takes up the question of what characteristics of humanity do we value enough to want to image them in AI and what are the consequences of those choices.[15]Herzfeld identifies three categories of ways scholars think about humans being made in the image of God. These categories are described as substantive, functional, and relational. The substantive approach to the Image of God places the emphasis or the locus on an aspect of human reason. For Augustine it was the will, for Reinhold Niebuhr it was the capacity for self-transcendence.[16] Regardless of the specific answer, those who fall into approach believe there is something particular to humans that we only share with God.[17] Critics of the substantive approach argue that there are few, if any, characteristics that

[15] Noreen L. Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, Theology and the Sciences (Minneapolis, MN: Fortress Press, 2002), 5.
[16] Ibid., 16.
[17] Ibid.

humans do not share with other species. Self-consciousness is found in elephants, dolphins, and some primates. Genetically speaking, humans share over 98% of their DNA with their closest related species. Even if there is something unique to humans, the substantive approach is also criticized by being too anthropocentric.[18]

Functional approaches to the Image of God place the locus in humanity's purpose. Scripturally, God gave humans dominion over the Earth, it is our responsibility to care for creation as representatives of God.[19] Critics of this approach argue that scripture does not provide a model for representing God that differs greatly from other civilizations of that time, and that those civilizations did not extend any relationship to the divine to all people, but rather only to the king or ruler.[20] Others would argue that humans have failed to live up to this function and therefore it cannot serve as the basis for the Image of God within humanity. The third approach, the relational interpretation, begins with the importance of relationality within God. The Trinity emphasizes the importance of relationality; it is a very part of the nature of God. Humans possess the same I-Thou relationship that God possess, but some scholars extend relationality further. Barth argues that the creation of humans as male and female is an extension of the relationality found in the Triune God.[21] Some criticize Barth because humans are far from the only species to have two sexes, while others argue that relationality is not the foundation of the Image of God, but rather a consequence of the Image, which is grounded either substantively or functionally.[22]

---

[18] Oliver Putz argues for an expanding of the Image of God to include other species.
[19] Herzfeld, *In Our Image*, 23-24.
[20] Ibid., 25.
[21] Ibid., 26.
[22] Ibid., 30.

Herzfeld finds a striking similarity between the ways in which intelligence in modeled in the context of AI and how theologians describe the image of God. Just as she provides three categories for the image of God, she therefore offers three similar categories for intelligence. Intelligence is seen as a quality[23], matching a function of human intelligence, or being able to relate to humans. Identifying the intelligence in AI as a quality, researchers believed that intelligence could be isolated and reproduced in mechanical structures. Referred to as symbolic AI, Herzfeld cites that this was the dominant approach on the field until the 1980s. The idea is that rational thought is made up of symbols, of which there are a fixed number used in combination with one another. Additionally, there are rules that govern the combinations and there must be a fixed number of these as well. Symbolic AI that have been produced are rather specialized in their knowledge.[24]

There is a divide in the AI community between strong AI and weak AI. Strong AI wants to replicate the full effects of the human mind in computers. Weak AI, on the other hand, focuses on one aspect of human intelligence, a particular function, and seeks to create AI that can do one task very well. Perhaps the most famous example of such a functional approach is Deep Blue, the chess playing computer.[25] It did not exhibit behavior that humans do in competition; it simply used the rules of the game and used its superior ability process moves ahead to overpower, in a sense, its human opponent.

The relational understanding of AI comes from the famous Turing Test. The abstract question of whether computers can think was impossible to answer without interacting with the

[23] Ibid., 35-49.
[24] Ibid., 38.
[25] Ibid., 43.

computer. Alan Turing posited an experiment where a person is connected via computer to two subjects, and asks them questions, with the intent to determine which is a human and which is a machine. An AI that could fool the interrogator as often as it failed would "pass" the test and could be described as intelligent.[26]

Jahoda's Latent Functions in the Context of AI

Depending on which kind of AI one is talking about, questions can be raised as to how this kind of intelligence is similar to humans and how Jahoda's latent functions could apply to nonhuman workers. In this section, I examine each of Jahoda's latent functions using Herzfeld's typology of intelligence to raise questions about the status of AI.

Time structure

AI that is doing work will have enforced activity for sure, but will there be an understanding of structured time? If a machine is working nonstop is there really time structure? If intelligence is a substantive characteristic, and it is something that separates humans from other organisms, if AI possesses a substantive intellect as well, will they require time structure in the way we do? Will their minds require rest as ours do? If intelligence in AI is relational, should AI be given time to have social interactions, and since our intelligence is embodied, should we expect AI to be as well? Would a network connection be sufficient for AI socialization, or would it require physical presence as well?

Regular shared experiences

[26] Ibid., 45-46.

How do we define shared experience? For AI, would doing their own job, even if part of an assembly line be considered shared experience? If Deep Blue played chess with another Deep Blue, would they be sharing a chess match or competition, or simply acting out their designed purposes? If AI is relational, will they experience dissatisfaction if they are forced to work in isolation?

Information about personal identity

Will AI have the same relationship between employment and identity as humans do? It seems to me that functional AI could evolve a "strong" connection between the function they perform and a sense of identity. If AI is relational, is it possible for them to form a sense of personal identity without social interaction? If AI is substantive, will AI come to similar conclusions about themselves as humans do, given that intellect is one of the things that people associate with human identity?

A link with collective purpose

In what kind of collective purpose could AI engage? Is it possible to program collective purpose of AI as part of its overall purpose? For instance, as an intelligent agent, could the collective purpose of AI include the flourishing of intelligent life, or abiding by the categorical imperative, creating rules that apply universally to all rational agents? The rules commonly referred to as Asimov's laws of robotics focus on nonmaleficence, but the requirements of nonmaleficence are less stringent than the requirements of beneficence. If AI are programmed to act beneficently towards humans, or more generally to all intelligent agents, does this place a collective purpose within their own *telos*?

Enforced activity

Do humans have the right to enforce activity upon other intelligent agents? Even if AI is only functionally intelligent, is that sufficient for a preferred moral status? Enforced activity assumes that when one is not working, one has the choice to not be active. As stated in the time structure section, will AI be given the opportunity to have "down time," and if not, if enforced activity a latent function that helps shape identity, or is it a burden imposed on those powerless to argue against it?

Transhumanism and the Possibility of Second Class Citizens

I believe that the same arguments presented about membership in society and the specter of second class citizens surrounding AI and automation can also be made regarding transhumanism. Although there is significant diversity within the movement about how to improve humans, the overall goal is to take control of human evolution in order to move beyond the current limits of humanity. To my knowledge, there is not much discussion about forming their own society or leaving the planet in order to form their own society. In fact, there seems to be a somewhat naïve belief that everyone will want to take advantage of the improvements to humanity that they seek. Such universal embracement is unlikely, especially at first, so there will be significant time when humans and transhumans, in a variety of forms, will exist together in society. This co-existence, however, presents the same problems as co-existing with AI.

Although I am reluctant to appeal to literature and popular culture, there are far too many examples of how people envision the relationship between humans and beings with abilities surpassing humans. Particularly when there are few enhanced individuals compared to humans, the relationship is often cast in terms of the dangers to humanity. Those who are not human, or beyond human are seen as possessing abilities that are too dangerous for individuals to possess. Although one could imagine a scenario in which transhumanists intentionally try and improve

human moral capacities, it is difficult to see how that would currently be possible, and also hard to believe that every transhumanist would want to pursue such improvements.

On the other hand, the more accepted human modification becomes, the relationship between humans and those with enhanced abilities changes. In the film Gattaca, the protagonist is born into a world that has just started to utilize gene editing. His parents decide to have their first child trusting God, not their "local geneticist," forgoing any improvements to his genome. It turns out he has a heart condition and the doctors say he will have a shorter life span than the average child. His younger sibling, born only a few years later, is conceived using IVF and gene editing to ensure a healthy embryo. By the time the protagonist is an adult, society has become stratified by genetic potential. Only those who have superior genetic profiles are eligible for the best jobs, whereas unaltered humans are relegated to lower paying jobs. The main character illegally assumes the identity of an enhanced person who has been paralyzed, a practice that is stigmatized, with people referring to such pretenders as "borrowed ladders" or "de-gene-erates." The enhanced person provides biological samples for the tests employees are subjected to in order to ensure they are who they say they are, and in turn receives housing and money.

If transhumanists increase their intelligence, they could argue that they know what is best for both humans and transhumans. If they increase their strength sufficiently, they could argue they have the power and therefore should rule. If their lives are extended significantly beyond the human lifespan, they could argue policies have a longer lasting effect on them, so they should have ultimate say, or by living longer, they can consolidate capital and political power over time. We live in a time where America is significantly divided, where people already have more power than others, is it that hard to believe that furthering the divide between abilities will further social division?

Reimaging Herzfeld's Typology for Transhumanism

I believe that Herzfeld had an important insight when comparing discussions of the Image of God and models of intelligence in AI research. Though I do not think the comparison is quite as clear, in this section I attempt to apply the three fold typology she uses to transhumanism. Instead of asking what aspect of human nature will be imparted to computers made in our image, the question in the context of transhumanism is what about human nature do transhumanists seek to improve or replace? The first category of the typology is the most straightforward, substantive enhancements. These are specific traits or characteristics that humans possess that transhumanists want to enhance, increasing the depth or power of the capacity. It could also apply to characteristics that transhumans want to be incorporated in a new or post-human nature that humans do not currently possess.

The second category of the typology, the functional approach, is probably the most difficult of the three to apply to transhumanism. Functional here is related closer to the typology of the Image of God than it is intelligence in the context of AI. Though it is plausible that transhumanists would still want humans and post-humans to steward creation, I would argue the primary function transhumanists value is the role of taking greater control of human evolution. The motivations behind this role might differ, such as a desire for immortality, or to minimize the role of chance in evolution and the removal of harmful conditions from the human genome, but what agents are doing remains the same; intentional, direct changes to what is considered human with the purpose of improving the human condition, or creating a species that is an improvement.

The third category, the relational aspect, focuses on how humans or post-humans relate to their bodies and to the rest of creation. Although some might view this as a subset of substantive

enhancements, I argue that the emphasis here is on how individuals exist in relationship. Reinhold Niebuhr referred to the perfection of the creaturely aspect of human nature as fulfilling the natural law, which he defines as harmony within three sets of relationships: one's relationship with God, an ordered soul (relationship of the self within the self), and one's relationship with the rest of creation. Transhumanism offers the possibility to improve all three of these sets of relationships. Increasing empathy or enhancing spiritual experiences could bring one closer to God. Tattoos that identify changes in blood sugar, or instant biofeedback that one can use to control breathing, meditation, blood pressure, etc. improves one's relationship with one's body and self. Expanding the range of light that people can see, or enhancing other senses can make people more aware of how they impact the world around them.

Tentative Steps Forward

In this conclusion I will draw upon the questions raised in the paper so far to offer tentative steps forward. These steps include what could or should be included in the *telos* of AI, how society needs to prepare for AI, automation, and unemployment, as well as what might be done to minimize of eliminate the existence of second class citizens in the context of AI and transhumanism. As I stated in the section on collective purpose, I believe that the *telos* of AI needs to include the flourishing of humans as well. A possible way of doing this is to move from nonmaleficent language to the more stringent standards of beneficence towards humans, or intelligent agents in general. Just as the perfection of human nature for Niebuhr can include how humans relate to others, which is another aspect of human nature besides intelligence that could be given to AI.

If the relationship between humans and AI is of primary importance to AI, humans will also need to prioritize the relationship, or risk AI questioning why the relationship is

predominately one way. Depending on how one views the intelligence of AI, different ways of dealing with this situation exist. If AI is functionally intelligent, humans could still decide to treat AI the same way other intelligent beings are treated. If intelligence is substantive, then AI should not be forced to work nonstop, and be given the opportunity to rest and reflect like human agents do. And if intelligence is only identified in relationship, then AI must be given the space and time to interact with others in ways that are meaningful to them.

In order to keep AI from being treated as second class citizens, the previous paragraphs comments are a start, but are insufficient. If AI are performing the kind of work that humans or post-humans do not desire to do, then they must be given membership in society and have a voice in collective decision making. Although a subservient class of automated, emotionless workers might be appealing to some, human intellect is rarely experienced as completely disconnected from emotion. Spock on Star Trek is such a compelling character because humans find it hard to relate to intelligence void of emotion. If AI is truly like human with regards to intelligence, emotion could be present, but even if it is not, AI still deserve respect and protection for the work they do for the community.

With the rise of AI and automation, steps can be taken to ensure that unemployed humans do not become second class citizens either. The first step is to work to reduce the stigmatization of unemployment. There will be many skilled workers who will be replaced by nonhuman workers; it has more to do with safety or finances than it does the worth of individual employees. A significant increase in unemployment will also likely require government action. This could include forming a universal basic income, or legislating the regulation of automation with taxes or charges that benefit directly those whose jobs have been displaced. Finally, additional opportunities need to be made for people who are unemployed to engage in meaningful social

leisure activities. This could include local governments creating volunteer projects, sports leagues, or even expanding representative governance.

There are multiple ways to ensure that neither enhanced humans, post-humans, nor humans become second class citizens. The widest reaching of these would be to legislate genetic freedom, with caveats to public welfare and health. If technological modifications are safe, people should be free to choose whether or not to pursue particular enhancements, and not be penalized in the workplace of society for their choices. If there are concerns about the expense of enhancements, then one way to level the playing field would be to institute a lottery. Anyone seeking a particular enhancement could submit their name, and then people are randomly chosen to receive the enhancement at a sliding scale of cost. If the enhancements are meant to improve society or the species as a whole, then a lottery is one way to minimize the chances of a stratified society based on the enhanced haves and the nonenhanced have nots. Human, enhanced humans, and post-humans all need governmental representation if they all coexist in society. Taxation without representation led to the American Revolutionary War, and the mistreatment of AI, humans, enhanced humans, or post-humans could lead to similar consequences, but with far greater potential for global damage, given the changes in technology since the 1700s.

Bibliography

Herzfeld, Noreen L. *In Our Image: Artificial Intelligence and the Human Spirit*. Theology and the Sciences. Minneapolis, MN: Fortress Press, 2002.

Mossakowski, Krysia N. "The Influence of Past Unemployment Duration on Symptoms of Depression Among Young Women and Men in the United States." American Journal of Public Health 99, no. 10 (October 2009): 1826-32.

Walzer, Michael. Spheres of Justice: a Defense of Pluralism and Equality. Reprint ed. New York: Basic Books, 1984.

Waters, Lea E., and Kathleen A. Moore. "Reducing latent deprivation during unemployment: The role of meaningful leisure activity." Journal of Occupational and Organizational Psychology 75, no. 1 (March 2002): 15-32.