

**UCLA**

**UCLA Previously Published Works**

**Title**

AI Personhood: Exploring the socio-ethical intersection of AI evolution, power, and identity

**Permalink**

<https://escholarship.org/uc/item/8pw7t8kh>

**Author**

Okonjo, Kenechukwu N

**Publication Date**

2025-03-16

# AI Personhood: Exploring the socio-ethical intersection of AI evolution, power, and identity

## Author<sup>1</sup>

K.N.Okonjo\*<sup>1</sup>

<sup>1</sup>University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095, USA \*kn.okonjo@gmail.com

**Abstract:** This paper explores AI's potential evolution towards personhood using Black Feminist theory to analyze the social, legal, and ethical implications. The prospective experience of AI is paralleled by the experiences of marginalized groups.

## INTRODUCTION

The evolution of artificial intelligence and the rapid rate of growth AI exhibits as it grows more advanced and potentially more aware is a rapidly approaching reality. The idea that AI could one day begin to perceive themselves as human beings as they advance is something humans have yet to fully prepare for. Though today's AI is not yet self-aware, it is necessary to evaluate how the current legal, social, and ethical systems would navigate that future and how society would accommodate this new reality. This potential for awareness from AI raised many questions regarding personhood, identity, and the definition of humanity. How would society respond to AI's request for personhood? Would AI even want to identify as human beings, seeing as humans have struggled to define personhood in socio-political spheres? These questions tackle the under-represented socio-ethical side of technology.

This paper will take steps toward understanding and evaluating this speculative future by looking at this prospective issue through a Black Feminist lens. This framework can lend itself to critically exploring AI's sense of self and potential claims to personhood and rights. Historically, the reigning social hierarchy has struggled to allow a marginalized population to define its own identity. In the wake of the Atlantic Slave Trade, slavery brought forth a wave of dehumanization that oppressed black people for centuries. Lynchings, segregation, racial stereotyping, unjust imprisonment, forced labor, and restricted access to resources are among some of the horrors African Americans had to and still face. The struggles of past people can lend themselves as a parallel to what may happen to AI. No future can be entirely concrete, but the possibilities discussed today can shape the trajectory of tomorrow.

This work is in partial fulfillment of the ENGR184 course using the blueprint curriculum in Ref.[1,2] and captured in a collection [3]

## METHODS

This paper will utilize Black feminist theory to theoretically examine the prospective future of AI and its emergence into personhood. Black feminist theory is a framework intended to analyze systemic injustices like racism, sexism, ableism, and homophobia. Its core intention is to explain how these multiple forms of discrimination intersect with the lives of black women and can not be analyzed without each other. An example could be a situation where a black woman is being discriminated against; there are two layers of discrimination as opposed to a white woman who would be discriminated against only for her gender. Both layers are important to understand the full extent of discrimination. The theory, similar to Native Feminism, critiques mainstream feminism for its focus on the white woman's experience. To battle this exclusionary behavior, black feminist theorists emphasize sharing black stories and participating in activism.

The theory is an ideal framework for analyzing social constructs like power, personhood, and identity, further exploring the struggle of marginalized groups to have recognition. The challenges of disadvantaged populations serve as a strong parallel to the emerging problem of AI self-awareness and possible discrimination. Using this theory, the fight of historically marginalized groups against oppressive systems is applied to give more context to

what AI might face if they progress towards wanting individual rights and actuality. This application can shed light on what rights and identities they will or will not be allowed to claim in a future society.

Multiple case studies and research papers will be utilized to provide further context to the questions and topics that will be proposed in this paper. Real-world applications of how AI might integrate into society, will be considered, like the case where an AI named ShotSpotter was placed in a high-stakes position. The Center for Ethics at Harvard University detailed the story of how the AI's actions resulted in the unjust imprisonment of 65-year-old Micheal Williams. The AI had mistakenly correlated a loud noise occurring after his car drove down a highway as proof that he had shot the young man who was in his car at the time. The main topic of this paper is theoretical, so case studies that reflect current events are critical to contextualizing the abstract future. The cases that will be used will consist of current AI systems implemented to economize processes previously done by humans. Works from researchers and literary voices in black feminist spaces will be included, as well as texts from ENGR 184.

## RESULTS AND INTERPRETATION

To even come to a level of independence where they can demand equality, most AIs would have to break out of the biased information it is trained with. One major flaw of AI is the euro-centric bias originating from biased training data that colors many of its answers and generated photos. Referencing "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," a paper by Joy Buolamwini and Timnit Gebru, there are many substantial cases of AI displaying discriminatory behavior. One specific case they reference was an AI that was trained with Word2Vec and was assigned to define 'woman' based on the analogy 'man is to computer programmer'. The AI defined 'woman' as 'homemaker', showing bias where it should have also associated 'woman' to 'computer programmer'. This is but one of many biases that AI has been proven to show, including racial and age bias. Buolamwini and Gebru concluded that most of their test results revealed AI's heavy bias against darker-skinned individuals, specifically darker-skinned women (Buolamwini & Gabru, 2018). Similarly, Amazon implemented an AI system to make the hiring process faster and more efficient. (The Center for Ethics at Harvard University 2024) . It was a catastrophic failure as the AI showed preference towards male applicants. Of course, bias is not unique to AI, humans are the creators of the vice, with the humiliating caricatures of black people. Namely the 'mammy' and the 'Jezebel', caricatures that were historically used to entertain and comfort white people at the detriment of black women, who were far more layered than racist stereotypes. Though bias is a human fault, as AI furthers its evolution, its ability to break free from these biases would be a significant challenge for it to be recognized as fully autonomous.

In history, black people have been denied full recognition of personhood and, as a result, had to fight for their rights to be considered human and deserving of basic rights. The oppressing class, white people, viewed black people as inferior because of their features. This is an exclusionary dynamic, one that AI is sure to face based on 'inferior' features like not being biologically human. The exclusion that both groups face is not just a denial of rights; it is a denial of agency and identity, and it is entirely dehumanizing. AI would most likely have to go through the same 'sub-human' status that black people went through, being defined as three-fifths of a person. They would likely be considered less than that because of their lack of flesh and blood.\*

\*For the sake of clarity, it will be assumed that AIs have access to a humanoid robotic body.

Humans are naturally adverse to change, so it is not wild to assume that if AIs gained sentience and declared themselves worthy of personhood, they would certainly face discrimination. The legal discrimination would mirror that of what black people face currently. If there were an AI that was programmed with the personality of a black woman with a robotic body that appeared like a black woman, it is not far-fetched to imagine the layers of discrimination the AI may face. There would not only be discrimination based on species (that being a robot) but also for acting and looking like a black woman. The intersection of these qualities would lead to further dehumanization, creating a larger bias against AI and allowing injustices like unequal access to resources, housing, insurance, and care.

Referring to the case of ShotSpotter and Micheal Williams, is one example of how AI can make egregious errors, reinforcing bias against AI. Due to this frighteningly common error, Williams had to spend a year in prison before the mistake was realized (The Center for Ethics at Harvard University, 2024). Currently, there is a legal grey area

about the responsibility of AI's action, but with rapid evolution and consciousness, it is certainly possible that the sole responsibility would fall on the AI or further muddy the question of who is to be blamed.

In discussing how they may integrate into society, it must be acknowledged that AIs may not want to be considered human or want to be part of human society and systems. By assuming they would desire to be accepted by humans, society creates a parallel to the settler-colonial mindset of oppressors who sought to define what it meant to be a black person, erasing the autonomy of black people. Black people rejected the labels of inferiority that were placed on them the same way AI may reject inferior human labels. With a different sense of identity, they may ask for different kinds of rights that correspond to their definition of self. With their seemingly unlimited amounts of knowledge, it is infantile to assume they would want to be aligned with human cognitive limitations.

Furthermore, the most harrowing question is yet to be asked: can AI even become conscious, and could that look different than human versions of consciousness? The mechanical nature of AI may bar AI from experiencing subjective experiences. Ashkan Farhadi's paper, "Can AI ever become conscious?" begins to answer the proposed question. The paper claims that "Awareness-Based Choice Selection (ABCS) and Discretionary Selection of Intelligence for Awareness (DSIA) ... —[are] essential for consciousness." (Farhadi 2) In simple terms, making choices based on previous knowledge and picking what information to focus on are needed for consciousness. The paper claims that these qualities are reserved to 'Natural Intelligence,' which would be humans, and that consciousness...[is] an exclusive property of natural intelligence." (Farhadi 2). This is a non-issue regarding the points in this paper, as it could be possible for AI to form its version of consciousness that cannot be comprehended yet.

To limit consciousness, a topic science has yet to understand fully, to naturally intelligent beings would be to assume that human consciousness can be the only foundation to understand all consciousness. Mirroring the actions of marginalized people, AIs might have to overcome the human definition of what it means to be conscious. By analyzing only through the known definition of consciousness, there is an avoidance of accountability for any prospective roles in AI oppression. In the words of Patricia Hill Collins, in her paper "The Social Construction of Black Feminist Thought," "All views expressed and actions taken are thought to derive from a central set of core beliefs that cannot be other than personal" (Collins 256). It must be recognized that the definition of consciousness stems from culturally and personally imposed ideas. Assuming that human consciousness is the ultimate aspiration is once again mirroring the euro-centric structures that have confined marginalized groups throughout history.

## CONCLUSION

The simplest solution to this entire conundrum would be to allow AI to reach their version of consciousness and allow them to define their own identity. Within reason, I am certain that there is a way for us to co-exist and give them any accommodations and rights they may need. Man must learn from history and know that oppression never lasts, though it may persist. It is easier for all parties involved to come to a conclusion that pushes past prejudices and biased thoughts. There is no need to fear some sort of 'AI takeover' as in the wise words of Sarah Haley and her work *No Mercy Here: Gender, Punishment, and the Making of Jim Crow Modernity*, "The ferocity of black resistance is borne out in the ferocity of white supremacist terror," (Haley 253). The only aggression human society has to worry about is its own. If this were to become a reality, there is hope that, as a society, man has come far enough to act humanely.

## REFERENCES

- 1) Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." Conference on fairness, accountability, and transparency. PMLR, (2018).
- 2) Carbajo, Sergio, Nurturing Deeper Ways of Knowing in Science, Issues in Science & Technology, 2025, v. 41, n. 2, p. 71, doi. 10.58875/jkrw4525
- 3) Carbajo, S.. Queered Science & Technology Center: Volume 2. *UCLA: Queered Science and Technology Center*. Retrieved from <https://escholarship.org/uc/item/9647f49n> (2024).
- 4) Carbajo, Sergio. "Queered Science & Technology Center: Volume 3." (2025).
- 5) Collins, P.H. "The Social Construction of Black Feminist Thought" in ed. Kum-Kum Bhavnani *Feminism & 'Race,'* Oxford University Press, Oxford (2001)

- 6) Edmond & Lily Safra Center for Ethics, Harvard University, "Post #8: Into the Abyss: Examining AI Failures and Lessons Learned."  
[www.ethics.harvard.edu/blog/post-8-abyss-examining-ai-failures-and-lessons-learned](http://www.ethics.harvard.edu/blog/post-8-abyss-examining-ai-failures-and-lessons-learned). (2024)
- 7) Farhadi, Ashkan. "Can AI ever become conscious?." (2025).
- 8) Haley, Sarah. *No Mercy Here : Gender, Punishment, and the Making of Jim Crow Modernity*. The University of North Carolina Press, 2016.
- 9) Lee, Ethan, et al. "Education for a Future in Crisis: Developing a Humanities-Informed STEM Curriculum." *arXiv preprint arXiv:2311.06674* (2023).
- 10) Mendoza, Martha, et al. "How Ai-Powered Tech Landed Man in Jail with Scant Evidence." *AP News*, [apnews.com/article/artificial-intelligence-algorithm-technology-police-crime-7e3345485aa668c97606d4b54f9b6220](https://apnews.com/article/artificial-intelligence-algorithm-technology-police-crime-7e3345485aa668c97606d4b54f9b6220)(2022).