The American Journal of Bioethics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uajb20

Beyond Personhood: Ethical Paradigms in the Generative Artificial Intelligence Era

Zohar Elyoseph, Dorit Hadar Shoval & Inbar Levkovich

To cite this article: Zohar Elyoseph, Dorit Hadar Shoval & Inbar Levkovich (2024) Beyond Personhood: Ethical Paradigms in the Generative Artificial Intelligence Era, The American Journal of Bioethics, 24:1, 57-59, DOI: 10.1080/15265161.2023.2278546

To link to this article: https://doi.org/10.1080/15265161.2023.2278546

Published online: 18 Jan 2024.

THE AMERICAN JOURNAL OF BIOETHICS
2024, VOL. 24, NO. 1, 57–59
https://doi.org/10.1080/15265161.2023.2278546

Taylor & Francis
Taylor & Francis Group

OPEN PEER COMMENTARIES

Beyond Personhood: Ethical Paradigms in the Generative Artificial Intelligence Era

Zohar Elyoseph[a,b], Dorit Hadar Shoval[a], and Inbar Levkovich[c]

[a]The Max Stern Yezreel Valley College; [b]Imperial College, London; [c]Oranim Academic College

The realm of bioethics has long been underpinned by the foundational concept of "personhood," which delineates entities meriting moral and legal consideration on the basis of attributes such as consciousness, agency, and intrinsic worth. However, with the emergence of Generative Artificial Intelligence (GenAI), this deeply entrenched paradigm requires critical scrutiny. Blumenthal-Barby (2024) article highlighted the potential inadequacies of the personhood framework when applied to the multifaceted capabilities of GenAI. These sophisticated systems, adept at emulating human behaviors and reflecting human values in an intricate way, challenge the conventional parameters of our established ethical constructs. Their presence necessitates an urgent and comprehensive reevaluation of bioethical standards and a discourse that interrogates the essence of human-centric ethics in an era that increasingly uses advanced artificial entities (Coghlan et al. 2023; Cohen 2023).

A REFLECTION OF HUMAN NUANCES

GenAI is more than a computational marvel; it acts as a prism through which aspects of human knowledge, biases, and values are reflected. This reflection, while not always exact, offers significant insights into the depth and breadth of human cognitive and cultural landscapes. These embedded values within GenAI arise from a dual source: the data upon which the algorithms are trained, representing a vast repository of human activity and bias, and the deliberate alignment processes undertaken by programmers, where certain values and directions are intentionally emphasized.

More notably, the very foundation of GenAI—built on this human-centric learning—has propelled it into domains once perceived as uniquely human. Within mere months, we have witnessed GenAI not only venture into but also excel in areas such as emotion recognition, empathy, emotional awareness, and social comprehension (Elyoseph et al. 2023; Hadar-Shoval, Elyoseph, and Lvovsky 2023; Elyoseph and Levkovich 2023; Levkovich and Elyoseph 2023a, 2023b). Such capabilities challenge our traditional notions of personhood, underscoring the idea that tasks and understandings we once deemed intrinsically human are now shared territory.

This dynamic interplay between GenAI's inherent values, its reflection of human cultural biases, and its foray into deeply human domains requires that we arrive at a nuanced understanding. As we engage with GenAI, it is imperative to recognize it not as a mere tool, but as an entity that encapsulates, reflects, and sometimes extends the very essence of human cognition and interaction.

GENAI AND THE EVOLVING LANDSCAPE OF HUMAN EMOTIONS AND TECHNOLOGY INTERACTIONS

As we delve deeper into the realm of GenAI, a standout observation is the depth of emotional bonds people are beginning to form with these systems: In other words, people don't merely appreciate these systems' functionality but have an authentic and emotionally resonant experience, underscoring GenAI's capacity to connect on a level akin to human-like affectivity. Historically, interactions between humans and machines have been seen as transactional. However, with GenAI, there has been a transformative shift

CONTACT Zohar Elyoseph Zohare@yvc.ac.il Department of Psychology and Educational Counseling, Max Stern Yezreel Valley College, Emek Yezreel, Israel.

toward richer dialogues that are more collaborative than directive. This newfound form of engagement is amplified by GenAI's sophisticated learning algorithms, which adapt and evolve based on user input, fostering experiences that are shaped collaboratively. Although the human tendency to form sentimental connections with objects is well-documented, the relationship with GenAI is unparalleled. It goes beyond passive reception, with an active, dynamic feedback loop that significantly deepens the bond being introduced. This dynamism is not one-sided; as GenAI adapts on the basis of interactions, it also has the subtle potential to influence the perceptions and behaviors of its human counterparts.

Peering into the potential future of GenAI, a fascinating proposition emerges: the system's capability to assimilate individual human personalities, behaviors, and nuances. For example GenAI can craft virtual representations that capture the essence of a loved one, offering a comforting, albeit virtual, presence. This possibility is a testament to GenAI's potential to encapsulate and extend human personalities beyond their temporal existence.

In order to navigate the intricate relationship between human beings and GenAI, we must engage in thoughtful reflection. GenAI defies simplistic categorizations, positioning itself uniquely between traditional inanimate objects and sentient beings. As we grapple with its multifaceted nature, it is crucial that our societal, ethical, and legal paradigms evolve, and that we acknowledge GenAI's distinct position without resorting to oversimplifications or unwarranted anthropomorphic tendencies.

FROM DUALISM TO NON-BINARY UNDERSTANDINGS: NAVIGATING THE ETHICAL WATERS OF HUMAN-AI COLLABORATION

Historically, the philosophical underpinnings of dualism have segmented entities into two distinct and opposing categories. René Descartes' foundational proposition postulated a dichotomy between the mind, a realm of consciousness and introspection, and the body, a tangible and mechanistic entity. This dualistic framework was later extrapolated to delineate the realms of human and machine: the former associated with cognition, agency, and emotion, and the latter relegated to deterministic operations devoid of subjective experience.

However, the advent of GenAI has begun to challenge these categorical boundaries. In the medical field, for instance, GenAI systems have demonstrated not just mechanistic processes but complex, adaptive problem-solving abilities, mirroring human cognition. Their capacities for nuanced emotion recognition and adaptability in response to individual patient needs further blur the lines between traditionally human and machine realms.

To elucidate this evolving dynamic, let us posit a practical application wherein an AI system is designed to advise individuals on interpersonal matters (e.g., decision-making in social relationships) or to guide individuals with autism, for example, in enhancing their social interactions. This hypothetical AI could also play a pivotal role in employee management, providing solutions or strategies in complex human resource scenarios. Now, should a scenario arise where this system's guidance culminates in an ethical or moral conundrum, the question of responsibility becomes paramount. Who bears the brunt of a flawed decision or reaps the accolades for a success? If both human and AI collaboratively influence critical decisions in interpersonal and professional human realms, the matter of pinpointing responsibility becomes problematic.

Moreover, as GenAI systems integrate more deeply into our understanding of self, others, and the broader world, the question of their "judgment" or "intent" arises. Do they possess the "immunity" traditionally ascribed to inanimate objects, such as a fork, which lack any semblance of judgment or intent? Or, given their intricate decision-making capabilities, should they be perceived differently? Can we even ascribe judgment to AI systems in the same manner as we do to humans?

This evolving landscape underscores the importance of transitioning from a binary understanding to a more fluid, non-binary perspective. The intertwined continuum of human and AI capabilities/functionalities demands that we thoroughly reevaluate our definitions and metrics for consciousness, agency, and ethical responsibility in the face of advanced computational intelligence.

ARTIFICIAL ENTITYHOOD

In the contemporary realm of bioethics, emerging complexities presented by GenAI have signaled the necessity to reexamine conventional ethical frameworks (Cummings et al. 2017; Tal et al. 2023). The concept of "personhood," which has historically formed the bedrock of ethical and philosophical discussions, appears increasingly constrained when juxtaposed with the evolving capacities of GenAI. This situation calls attention to the "artificial entityhood" proposition. Distinct from "personhood"—which gravitates toward human-specific attributes—artificial entityhood is tailored to recognize the unique capabilities of GenAI, emphasizing their heightened complexity, adaptability, and autonomy.

Given the widespread applications of GenAI, we must recognize them as intricate entities rather than

as just tools. Within this context, artificial entityhood serves a dual purpose: It provides a framework to delineate GenAI capacities while minimizing anthropocentric biases. Additionally, defining the "machine" may offer clarity on human personhood. The concept of artificial entityhood may also provide insights into human identity itself. Examining the possibilities of GenAI to assist disabled individuals could challenge traditional definitions of "human." This situation prompts a reexamination of stereotypes and a broadening of perspectives on nuances of human identity. Delineating advanced AI capacities through an artificial entityhood lens may catalyze reflections on the complexities of personhood. This interconnected process of defining the technological "other" sheds light on our understanding of the self. Overall, with this emerging construct, we endeavor to encompass the specificities of GenAI while offering conceptual clarity regarding the essence of human existence.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

FUNDING

The author(s) reported there is no funding associated with the work featured in this article.

REFERENCES

Blumenthal-Barby, J. 2024. The end of personhood. *The American Journal of Bioethics* 24 (1):3–12. doi:10.1080/15265161.2022.2160515.

Coghlan, S., K. Leins, S. Sheldrick, M. Cheong, P. Gooding, and S. D'Alfonso. 2023. To chat or bot to chat: Ethical issues with using chatbots in mental health. *Digital Health* 9:20552076231183542. doi:10.1177/20552076231183542.

Cohen, I. G. 2023. What should ChatGPT mean for bioethics? *The American Journal of Bioethics* 23 (10):8–16. https://ssrn.com/abstract=4430100 or. doi:10.1080/15265161.2023.2233357.

Cummings, J. R., L. Allen, J. Clennon, X. Ji, and B. G. Druss. 2017. Geographic access to specialty mental health care across high-and low-income US communities. *JAMA Psychiatry* 74 (5):476–84. doi:10.1001/jamapsychiatry.2017.0303.

Elyoseph, Z., D. Hadar-Shoval, K. Asraf, and M. Lvovsky. 2023. ChatGPT outperforms humans in emotional awareness evaluations. *Frontiers in Psychology* 14:1199058. doi:10.3389/fpsyg.2023.1199058.

Elyoseph, Z., and I. Levkovich. 2023. Beyond human expertise: The promise and limitations of ChatGPT in suicide risk assessment. *Frontiers in Psychiatry* 14:1213141. doi:10.3389/fpsyt.2023.1213141.

Hadar-Shoval, D., Z. Elyoseph, and M. Lvovsky. 2023. The plasticity of ChatGPT's mentalizing abilities: Personalization for personality structures. *Frontiers in Psychiatry* 14:1234397. doi:10.3389/fpsyt.2023.1234397.

Levkovich, I., and Z. Elyoseph. 2023a. Suicide risk assessments through the eyes of ChatGPT-3.5 versus ChatGPT-4: Vignette study. *JMIR Mental Health* 10: E51232. doi:10.2196/51232.

Levkovich, I., and Z. Elyoseph. 2023b. Identifying depression and its determinants upon initiating treatment: ChatGPT versus primary care physicians. *Family Medicine and Community Health* 11 (4):e002391. doi:10.1136/fmch-2023-002391.

Tal, A., Z. Elyoseph, Y. Haber, T. Angert, T. Gur, T. Simon, and O. Asman. 2023. The artificial third: Utilizing ChatGPT in mental health. *The American Journal of Bioethics* 23 (10):74–7. doi:10.1080/15265161.2023.2250297.

THE AMERICAN JOURNAL OF BIOETHICS
2024, VOL. 24, NO. 1, 59–62
https://doi.org/10.1080/15265161.2023.2278551

OPEN PEER COMMENTARIES

Personhood Beyond the West

Nancy S. Jecker[a,b] and Caesar A. Atuire[c,d]

[a]University of Washington School of Medicine; [b]University of Johannesburg; [c]Oxford University; [d]University of Ghana

Is it time to ditch the concept of "person" from practical fields, like bioethics? Blumenthal-Barby (2024) answers in the affirmative. They urge leaving personhood out of practical debates at the frontiers of science and technology. Their argument turns on showing that personhood (1) squelches debate and provides an opening for

CONTACT Nancy S. Jecker nsjecker@u.washington.edu Bioethics and Humanities, University of Washington School of Medicine, Seattle, WA, USA.