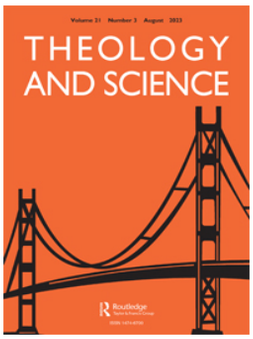

Theology and Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rtas20

What Does it Mean to Consider AI a Person?

Mark Graves

To cite this article: Mark Graves (2023) What Does it Mean to Consider AI a Person?, Theology and Science, 21:3, 348-353, DOI: 10.1080/14746700.2023.2230424

To link to this article: https://doi.org/10.1080/14746700.2023.2230424

THEOLOGY AND SCIENCE
2023, VOL. 21, NO. 3, 348–353
https://doi.org/10.1080/14746700.2023.2230424

CTNS
Routledge
Taylor & Francis Group

EDITORIAL

What Does it Mean to Consider AI a Person?

As artificial intelligence (AI) further pervades society, it raises a number of ethical issues as well as optimistic and pessimistic expectations for its effect. Theologians and religious ethicists can and should bring the wisdom of the world's religions to the immediate conversation.[1] Regardless of further major AI research breakthroughs, the impact of current AI technology on work, entertainment, political discourse, and other aspects of society will be substantial, especially given the immense corporate resources currently dedicated to applying that technology. AI advances will impact theology,[2] and the present editorial proposes one way theology can constructively impact AI.

At the core of speculation on AI sentience, consciousness, moral responsibility, agency, and its possible intentions toward humanity is a question of whether, or to what extent, we should consider AI as a person. Although historical theological, scientific, and ethical theories of personhood influence contemporary discourse about AI, most technologists lack the religious literacy to identify these theories' historical roots and the scholarly skills to reevaluate them in the current context. The significance of personhood is also exacerbated by human tendency to anthropomorphize.[3] Theologians, engaging contemporary science, can characterize what it would mean for AI to be a person, informing contemporary conversations and clarifying the imagination of those AI developers who attempt to integrate cognitive, social, and ethical aspects of AI in ways analogous to a person, especially for Artificial General Intelligence (AGI).[4]

Theologians and others have previously examined issues in AI and theology, and these efforts have built upon two main paradigms, or research programs, in AI: symbolic or Good Old-Fashioned AI and subsymbolic or statistical machine learning approaches, including the deep learning underneath the current explosion of technologies built upon GPT and other foundation models.[5] Considering how theologians have engaged each AI paradigm lays a foundation for developing a new AI and theology research program that may inform future AI development, instead of merely reacting to it.

Elsewhere I argue a plausible, near future advance in AI may arise from the synthesis of empirically oriented statistical machine learning mimicking perceptual processes and rationally grounded symbolic AI mimicking deliberative cognitive processes.[6] The construction of the next generation of AI may directly depend upon integrating aspects of what is usually considered unique to human persons, and theologians can clarify those theories and prepare resources for the subsequent public discourse. Regardless of whether an AI architecture meets a particular definition of personhood, or whether it requires time for intermediate advances, clarifying the possibilities of personhood will certainly be needed for meaningful public discourse given the anthropomorphizing already occurring with AI. As a step toward these efforts, I briefly review existing work in theology centered around aspects of personhood, such as, *imago Dei,* theological anthropology, and morality. Although the review is certainly not exhaustive, it hopefully represents the field sufficiently to orient new scholars to the area.

Initial theological engagements with symbolic AI include Ian Barbour's exploration of AI's relevance for human nature and Noreen Herzfeld's examination of the relationship between *imago Dei* and the human desire to build artificial intelligent systems in our image.[7] Anne Foerst used her work in robotics to highlight conflicts between work in artificial intelligence and Christianity.[8] William Clocksin examined AI with a focus on the role of social relationships in the intelligent behavior of AI as part of a social group, and Robert Geraci compared religious roots of the US focus on AI personhood with Japanese efforts to build humanoid robots.[9] Russell Bjork focused on questions of emergence and personhood within his exploration of AI and soul from historical and biblical perspectives; Andrew Porter dove into Heidegger's and Kierkegaard's understanding of the self, as it relates to AI; and Dion Forster examined identity and relationality in AI from an African perspective.[10] Each of these researchers touched on aspects of AI and personhood relevant to contemporary investigation, though they were limited to the predominant core theories of the time, namely those of symbolic AI.

Theologians and ethicists have engaged with statistical approaches to AI, including contemporary robotics. Theology journals have devoted special issues to AI and philosophy of religion, moral theology, transhumanism, and apocalypticism.[11] Scholars have examined AI soteriology,[12] AI as pastoral caregivers,[13] religious AI,[14] and AI contributions to theological inquiry.[15] Philip Hefner has extended his understanding of human nature as created co-creator to explore what it would mean to create AI co-creators.[16] Jordan Wales investigates AI through the lens of Augustine's theological understandings of creation as *rationes seminales* (seedlike principles) and memory; Robert Dell'Oro explores personhood in the context of Levinas's phenomenological account of the other; and Michael Burdett examines personhood from Buber's "I-You" perspective.[17] Several theologians and ethicists have examined AI in the context of transhumanism and human enhancement as it relates to personhood, morality, social inequity, and *theosis*.[18] Dina Babushkina and Athanasios Votsis argue that artificial identity in human-machine interaction is a better ethical focus than personhood, and I examine AI development of a self in community.[19] Simon Balle explores theological dimensions of humanlike robots in terms of anthropology, eschatology, ethics, and religious practices.[20] Anna Puzio and others have edited a volume on AI technology and theology that includes contributions on theological anthropology (in English and in German).[21] *Imago Dei* is a common point of engagement for many of these scholars, including ongoing work by Herzfeld.[22] Sara Lumbreras uses *imago Dei* as a starting point to examine artificial consciousness, experience, and subjectivity, and Marius Dorobantu argues for relationality as key to human-level AI.[23] Others have examined personhood in the legal context of synthetic persons, legal subjects, and granting of human rights to robots.[24] Hadi Akbar Dahlan suggests questions from Islamic thought to consider with respect to AI, and Alif Nawi *et al.* survey Muslim experts perspectives on AI.[25] CS Wareham and Chammah Judex Kaunda each examine AI as person from African accounts of personhood and relationality, respectively; Geraci examines AI in India; and Neela Bhattacharya Saxena takes a Buddhist perspective on impermanence to argue against transhumanist and Western assumptions of the person.[26]

As AI developers continue to use psychological, social, moral, and spiritual dimensions of the human person for imitating human thinking and as targets for modeling human rationality,[27] a clearer understanding of an AI person capable of moral and practical reasoning and aware of human spiritual strivings can lead to a beneficial and flourishing human future. Given the potential need to use AI to monitor the large-scale, opaque, and quickly occurring ethical impacts of other AI, attention to the moral dimension of an AI person is particularly crucial. As mentioned, several scholars have found a relational interpretation of *imago Dei* useful, and that appears a valuable Christian foundation, especially when it emphasizes

that AI exists in a social context comprised of relationships, not just on an isolated machine. The themes of identity, memory, and embodiment can build upon this foundation, and integrating their theological investigations with associated psychological theories may strengthen their application. Philip Hefner's recognition of creation and creativity also appears highly relevant for considering possible AI personhood as well as its construction. Non-Christian perspectives are important, especially for global AI development, and religious practices can be significant for those conversations (and ecumenical Christian ones). A focus on what it means for AI to be a person can orient ongoing AI development and help provide a framework for discussing its implications and considering what kind of AI person ought to be constructed. If humanity's future includes AI people, do we hope they are efficient, docile, obedient, virtuous, righteous, compassionate, or merciful? Considering what it means for AI to be a person generates important theological questions and can have significant positive impact on a major social change.

Notes

1. This is an explicit goal of the AI & Faith community of experts. https://aiandfaith.org/
2. Mark Graves, "ChatGPT's Significance for Theology," *Theology and Science* 21, no. 2 (2023): 201–4, https://doi.org/10.1080/14746700.2023.2188366.
3. Nicholas Epley, Adam Waytz, and John T. Cacioppo, "On Seeing Human: A Three-Factor Theory of Anthropomorphism," *Psychological Review* 114, no. 4 (October 2007): 864–86, https://doi.org/10.1037/0033-295X.114.4.864; Luisa Damiano and Paul Dumouchel, "Anthropomorphism in Human–Robot Co-Evolution," *Frontiers in Psychology* 9 (2018), https://doi.org/10.3389/fpsyg.2018.00468; Indrit Troshani et al., "Do We Trust in AI? Role of Anthropomorphism and Intelligence," *Journal of Computer Information Systems* 61, no. 5 (2021): 481–91, https://doi.org/10.1080/08874417.2020.1788473; Arleen Salles, Kathinka Evers, and Michele Farisco, "Anthropomorphism in AI," *AJOB Neuroscience* 11, no. 2 (2020): 88–95, https://doi.org/10.1080/21507740.2020.1740350; Amani Alabed, Ana Javornik, and Diana Gregory-Smith, "AI Anthropomorphism and Its Effect on Users' Self-Congruence and Self–AI Integration: A Theoretical Framework and Research Agenda," *Technological Forecasting and Social Change* 182 (2022): 121786, https://doi.org/10.1016/j.techfore.2022.121786.
4. Scott McLean et al., "The Risks Associated with Artificial General Intelligence: A Systematic Review," *Journal of Experimental & Theoretical Artificial Intelligence*, 2021, 1–15, https://doi.org/10.1080/0952813X.2021.1964003; Nanyi Fei et al., "Towards Artificial General Intelligence via a Multimodal Foundation Model," *Nature Communications* 13, no. 1 (2022): 3094, https://doi.org/10.1038/s41467-022-30761-2.
5. Brain Cantwell Smith and others refer to symbolic AI and statistical machine learning as two waves, with DARPA proposing goals for a third wave. However, considering the research programs as "waves" glosses over their parallel development and interspersed "winters" and omits the framework to synthesize them into a third program through their distinct but overlapping "cores." Symbolic AI had winters during mid-to-late 1970s, initiated by the Lighthill report's criticism of AI's intractable scalability, and thus impracticality; and from mid-1980s to mid-1990s with John McCarthy's critique of expert systems an early initiator. Statistical machine learning arguably began with Rosenblatt's development of perceptrons, stalled with Minsky and Pappert's critical book *Perceptrons*, then warmed again with backpropagation and parallel distributed processing (PDP), waiting until data quantity and computer power could take advantage of the approach to expand further. Nancey Murphy, "Theology and Science within a Lakatosian Program," *Zygon* 34, no. 4 (1999): 629–42, https://doi.org/10.1111/0591-2385.00241; John Haugeland, *Artificial Intelligence: The Very Idea* (Cambridge, MA: MIT Press, 1985). Brian Cantwell Smith, *The Promise of Artificial Intelligence: Reckoning and Judgment* (Cambridge, MA: The MIT Press, 2019);

John Launchbury, "A DARPA Perspective on Artificial Intelligence," *Machine Learning* (blog), March 19, 2017, https://machinelearning.technicacuriosa.com/2017/03/19/a-darpa-perspective-on-artificial-intelligence/; James Lighthill, "Artificial Intelligence: A General Survey," in *Artificial Intelligence: A Paper Symposium* (Science Research Council London, 1973), 1–21, http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm; F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review* 65, no. 6 (1958): 386–408, https://doi.org/10.1037/h0042519; Marvin Minsky and Seymour A. Papert, *Perceptrons: An Introduction to Computational Geometry* (Cambridge, MA: MIT Press, 1969); John McCarthy, "Some Expert Systems Need Common Sense," in *Proc. of a Symposium on Computer Culture: The Scientific, Intellectual, and Social Impact of the Computer*, vol. 426 (New York Academy of Sciences, 1984), 129–37, https://doi.org/10.1111/j.1749-6632.1984.tb16516.x; David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, "Learning Representations by Back-Propagating Errors," *Nature* 323, no. 6088 (1986): 533–36; David E. Rumelhart and James L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations, Psychological and Biological Models* (Cambridge, MA: MIT Press, 1987); Terrence J. Sejnowski, "The Unreasonable Effectiveness of Deep Learning in Artificial Intelligence," *Proceedings of the National Academy of Sciences of the United States of America* 117, no. 48 (2020): 30033–38, https://doi.org/10.1073/pnas.1907373117; Rishi Bommasani et al., "On the Opportunities and Risks of Foundation Models" (arXiv, July 12, 2022), https://doi.org/10.48550/arXiv.2108.07258.

6. Mark Graves, "Embodied Experience in Socially Participatory Artificial Intelligence," *Zygon*, 2023, https://doi.org/10.1111/zygo.12910; Mark Graves, "Theological Foundations for Moral Artificial Intelligence," *Journal of Moral Theology* 11, no. Special Issue 1 (March 2022): 182–211, https://doi.org/10.55476/001c.34130. See also, Marta Garnelo and Murray Shanahan, "Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations," *Current Opinion in Behavioral Sciences*, SI: 29: Artificial Intelligence (2019), 29 (October 1, 2019): 17–23, https://doi.org/10.1016/j.cobeha.2018.12.010.
7. Ian G Barbour, "Neuroscience, Artificial Intelligence, and Human Nature: Theological and Philosophical Reflections," *Zygon* 34, no. 3 (1999): 361–98, https://doi.org/10.1111/0591-2385.00222; Noreen L Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit* (Minneapolis, MN: Fortress Press, 2002).
8. Anne Foerst, *God in the Machine: What Robots Teach Us about Humanity and God* (New York: Dutton, 2004).
9. William F. Clocksin, "Artificial Intelligence and the Future," *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 361, no. 1809 (2003): 1721–48, https://doi.org/10.1098/rsta.2003.1232; Robert M Geraci, "Spiritual Robots: Religion and Our Scientific View of the Natural World," *Theology and Science* 4, no. 3 (2006): 229–46, https://doi.org/10.1080/14746700600952993.
10. Russell C. Bjork, "Artificial Intelligence and the Soul," *Perspectives on Science and Christian Faith* 60, no. 2 (2008): 95–102; Andrew Peabody Porter, "A Theologian Looks at AI," in *2014 AAAI Fall Symposium Series*, 2014; Dion Angus Forster, "Validation of Individual Consciousness in Strong Artificial Intelligence: An African Theological Contribution" (PhD Thesis, University of South Africa Pretoria, 2006).
11. Andrea Vestrucci, "Introduction: Five Steps Toward a Religion–AI Dialogue," *Zygon*, 2023, https://doi.org/10.1111/zygo.12828; Matthew J. Gaudet, "An Introduction to the Ethics of Artificial Intelligence," *Journal of Moral Theology* 11, no. Special Issue 1 (2022), https://doi.org/10.55476/001c.34121; Ted Peters, "Theologians Testing Transhumanism," *Theology and Science* 13, no. 2 (2015): 130–49, https://doi.org/10.1080/14746700.2015.1023524; Robert M. Geraci and Simon Robinson, "Introduction to the Symposium on Artificial Intelligence and Apocalypticism," *Zygon* 54, no. 1 (2019): 149–55, https://doi.org/10.1111/zygo.12489.
12. Eugene A. Curry, "Artificial Intelligence and Baptism: Cutting a Gordian Knot," *Theology and Science* 20, no. 2 (2022): 156–65, https://doi.org/10.1080/14746700.2022.2051248.

13. William Young, "Reverend Robot: Automation and Clergy," *Zygon* 54, no. 2 (2019): 479–500, https://doi.org/10.1111/zygo.12515; William Young, "Virtual Pastor: Virtualization, AI, and Pastoral Care," *Theology and Science* 20, no. 1 (2022): 6–22, https://doi.org/10.1080/14746700.2021.2012915.
14. Yong Sup Song, "Religious AI as an Option to the Risks of Superintelligence: A Protestant Theological Perspective," *Theology and Science* 19, no. 1 (2021): 65–78, https://doi.org/10.1080/14746700.2020.1825196.
15. Marius Dorobantu, "Artificial Intelligence as a Testing Ground for Key Theological Questions," *Zygon*, 2023, https://doi.org/10.1111/zygo.12831; Mark Graves, "AI Reading Theology: Promises and Perils," in *AI and IA: Utopia or Extinction?*, vol. 5, Agathon Journal (ATF Press, 2018); Mark Graves, "Computational Topic Models for Theological Investigations," *Theology and Science* 20, no. 1 (2022), https://doi.org/10.1080/14746700.2021.2012922; Andrea Vestrucci, "Artificial Intelligence and God's Existence: Connecting Philosophy of Religion and Computation," *Zygon*, 2023, https://doi.org/10.1111/zygo.12829.
16. Philip Hefner et al., *Human Becoming in an Age of Science, Technology, and Faith*, ed. Jason P. Roberts and Mladen Turk (Lanham: Fortress Academic, 2022).
17. Jordan Joseph Wales, "Metaphysics, Meaning, and Morality: A Theological Reflection on AI," *Journal of Moral Theology* 11, no. Special Issue 1 (2022): 157–81, https://doi.org/10.55476/001c.34129. Roberto Dell'Oro, "Can a Robot Be a Person? De-Facing Personhood and Finding It Again with Lévinas," *Journal of Moral Theology* 11, no. Special Issue 1 (2022): 132–56, https://doi.org/10.55476/001c.34128; Michael S. Burdett, "Personhood and Creation in an Age of Robots and AI: Can We Say 'You' to Artifacts?," *Zygon* 55, no. 2 (2020): 347–60, https://doi.org/10.1111/zygo.12595.
18. Noreen Herzfeld, "More than Information: A Christian Critique of a New Dualism," *Theology and Science* 14, no. 1 (2016): 84–92, https://doi.org/10.1080/14746700.2015.1122337; Braden Molhoek, "Sensuality and Altering Anthropology in Artificial Intelligence and Transhumanism," *Theology and Science* 14, no. 1 (2016): 99–104, https://doi.org/10.1080/14746700.2015.1122328; Ilia Delio, "Transhumanism or Ultrahumanism? Teilhard de Chardin on Technology, Religion and Evolution," *Theology and Science* 10, no. 2 (May 1, 2012): 153–66, https://doi.org/10.1080/14746700.2012.669948. Adam M. Willows, "Supplementing Virtue: The Case for a Limited Theological Transhumanism," *Theology and Science* 15, no. 2 (2017): 177–87, https://doi.org/10.1080/14746700.2017.1299375; Brian Patrick Green, "Transhumanism and Roman Catholicism: Imagined and Real Tensions," *Theology and Science* 13, no. 2 (2015): 187–201, https://doi.org/10.1080/14746700.2015.1023528. I Sil Yoon, "Amartya Sen's Capabilities Approach: Resistance and Transformative Power in the Age of Transhumanism," *Zygon* 56, no. 4 (2021): 874–97, https://doi.org/10.1111/zygo.12740. Ron Cole-Turner, "Theosis and Human Enhancement," *Theology and Science* 16, no. 3 (2018): 330–42, https://doi.org/10.1080/14746700.2018.1488526.
19. Dina Babushkina and Athanasios Votsis, "Disruption, Technology and the Question of (Artificial) Identity," *AI and Ethics* 2, no. 4 (2022): 611–22, https://doi.org/10.1007/s43681-021-00110-y; Mark Graves, "Shared Moral and Spiritual Development Among Human Persons and Artificially Intelligent Agents," *Theology and Science* 15, no. 3 (2017): 333–51, https://doi.org/10.1080/14746700.2017.1335066; Graves, "Theological Foundations for Moral Artificial Intelligence."
20. Simon Balle, "Theological Dimensions of Humanlike Robots: A Roadmap for Theological Inquiry," *Theology and Science* 21, no. 1 (2023): 132–56, https://doi.org/10.1080/14746700.2022.2155916.
21. Anna Puzio, Nicole Kunkel, and Hendrik Klinge, *Alexa, wie hast du's mit der Religion?* (DE: wbg, 2023), https://doi.org/10.53186/1030373. This volume also includes Katherine Schmidt, "Learn, Remember, Act: Theological Anthropology and AI Metaphor."
22. Noreen Herzfeld, *The Artifice of Intelligence: Divine and Human Relationship in a Robotic Age* (Minneapolis, MN: Fortress Press, 2023).
23. Sara Lumbreras, "Strong Artificial Intelligence and Imago Hominis: The Risks of a Reductionist Definition of Human Nature," in *Issues in Science and Theology: Are We Special?*

Human Uniqueness in Science and Theology, ed. Michael Fuller et al., Issues in Science and Religion: Publications of the European Society for the Study of Science and Theology (Cham: Springer International Publishing, 2017), 157–68, https://doi.org/10.1007/978-3-319-62124-1_11; Sara Lumbreras, "Lessons from the Quest for Artificial Consciousness: The Emergence Criterion, Insight-Oriented AI, and Imago Dei," *Zygon*, 2023, https://doi.org/10.1111/zygo.12827. Marius Dorobantu, "Human-Level, but Non-Humanlike: Artificial Intelligence and a Multi-Level Relational Interpretation of the Imago Dei," *Philosophy, Theology and the Sciences* 8, no. 1 (2021): 81–107, https://doi.org/10.1628/ptsc-2021-0006; Marius Dorobantu, "Imago Dei in the Age of Artificial Intelligence: Challenges and Opportunities for a Science-Engaged Theology – The ISCAST Journal," *Christian Perspectives on Science and Technology*, New Series, 1 (2023): 175–96, https://doi.org/10.58913/KWUU3009.

24. Robert M. Geraci, *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality* (Oxford: Oxford University Press, 2010), chap. 4; Sylwia Wojtczak, "Endowing Artificial Intelligence with Legal Subjectivity," *AI & SOCIETY* 37, no. 1 (2022): 205–13, https://doi.org/10.1007/s00146-021-01147-7; John-Stewart Gordon and Ausrine Pasvenskiene, "Human Rights for Robots? A Literature Review," *AI and Ethics* 1, no. 4 (2021): 579–91, https://doi.org/10.1007/s43681-021-00050-7.
25. Hadi Akbar Dahlan, "Future Interaction between Man and Robots from Islamic Perspective," *International Journal of Islamic Thought* 13, no. 1 (2018), https://doi.org/10.24035/ijit.06.2018.005; Aliff Nawi et al., "A Preliminary Survey of Muslim Experts' Views on Artificial Intelligence," *Islāmiyyāt* 43, no. 2 (2021): 3–16, https://doi.org/10.17576/islamiyyat-2021-4302-01.
26. C. S. Wareham, "Artificial Intelligence and African Conceptions of Personhood," *Ethics and Information Technology* 23, no. 2 (June 2021): 127–36, https://doi.org/10.1007/s10676-020-09541-3; Chammah Judex Kaunda, "Bemba Mystico-Relationality and the Possibility of Artificial General Intelligence (AGI) Participation in Imago Dei," *Zygon* 55, no. 2 (2020): 327–43, https://doi.org/10.1111/zygo.12598; Robert M. Geraci, *Futures of Artificial Intelligence: Perspectives from India and the U.S.* (Delhi: Oxford University Press, 2022); Neela Bhattacharya Saxena, "AI as Awakened Intelligence: Buddha, Kurzweil and the Film Her," *Theology and Science* 18, no. 1 (2020): 74–85, https://doi.org/10.1080/14746700.2019.1710351.
27. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th edition (Hoboken: Pearson, 2020), chap. 1.

Notes on Contributor

Mark Graves is Research Fellow and Director at AI & Faith and Research Associate Professor of Psychology at Fuller Theological Seminary. He holds a PhD in computer science; has completed fellowships in genomics, moral psychology, and moral theology; and has published over seventy technical and scholarly works in computer science, biology, psychology, and theology, including three books.

Mark Graves
http://orcid.org/0000-0003-3146-3203