

THE BIG QUESTION

How much like us?

Lawrence A. Whitney 

Center for Mind and Culture, Boston, MA, USA

Email: lawrence.whitney.lc@gmail.com

(Received 1 April 2025; revised 2 April 2025; accepted 3 April 2025)

Abstract

The recent phenomenon of anthropomorphizing artificial intelligences (AIs) is uniquely provocative for philosophy of religion because of its tendency to place AIs in an analogous position to divinity vis-à-vis humans in spite of AIs being human artefacts. In the case of divinity, intelligent mental capacities are, and in the case of AIs are sometimes presented as inevitably becoming, not just equivalent but in fact superior to their realization in humans. Philosophers of religion would do well to learn from discussions of anthropomorphism in AI, in conversation with the historical debates over anthropomorphizing divinity, and to remember that evolved cognitive biases may lose their adaptive functions as the cultural context shifts, and even become maladaptive.

Keywords: anthropomorphism; AI; intelligence; consciousness; cognitive bias

Anthropomorphism has been a recognized problem in philosophy of religion since at least Xenophanes (Leshner 2023). Awareness of the human propensity to ascribe human traits and capacities to non-human animals, natural entities, and human artefacts is likewise long-standing. Yet the recent phenomenon of anthropomorphizing artificial intelligences (AIs) is uniquely provocative for philosophy of religion because of its tendency to place AIs in an analogous position to divinity vis-à-vis humans in spite of AIs being human artifacts. Anthropomorphism of AIs may seem inevitable given that the human capacity of intelligence is imposed on their computational capacities in the very name ‘artificial intelligence’, but so is their status as human artefacts. Core to the reshaping of philosophy of religion that AIs provoke, then, is the claim that humans have made, or are on the verge of making, the same superior intelligence uniquely ascribed to divinity in much philosophy of religion.

Philosophers of religion dispute whether intelligence or any other human attribute should be ascribed to divinity, and if so the degree to which it is in any way comparable to the human version. Some insist on the absolute aseity of divinity in order to maintain conceptual coherence, while others admit varying degrees of anthropomorphism in order for divinity to have relevance for human affairs. A parallel range of positions has emerged with regard to AIs, with some researchers insisting that ascribing intelligence to these computational systems is a category error even as others foresee them – especially Large Language Models (LLMs) – as exhibiting agency, intelligence, and either already or on the verge of consciousness and sentience (Melanie and Krakauer 2023). Notably, with respect to both divinity and AIs, it is human mental capacities such as intelligence, consciousness, and so on, that are being ascribed anthropomorphically; any moral attributes ascribed would be

downstream therefrom. In the case of divinity these mental capacities are, and in the case of AIs are sometimes presented as inevitably becoming, not just equivalent but in fact superior to their realization in humans. Since both divinity and AIs are black boxes, their inner workings are obscured so as to make explanation of how they fit these mental categories difficult if not impossible. What is surprising, then, is that philosophers of religion have yet to acknowledge, let alone address, that the same anthropomorphism that perennially arises with respect to divinity is now divinizing AIs.

AI researchers have developed a conceptual framework with respect to anthropomorphizing AIs that is helpful for philosophers of religion to understand the religious dynamics of anthropomorphism with respect to divinity: the Uncanny Valley. The phenomenon was first hypothesized in 1970 by Masahiro Mori and describes the non-linear affective trajectory of human experience of increasingly anthropomorphized robots, and presumably other computational systems (Mori 2012; Shensheng et al. 2015). The idea is that human positive affective response increases proportional to degree of anthropomorphism to a point, after which it craters, even turning negative, until the degree closely approximates full humanity, at which point positive affective response skyrockets again. The implication is that an AI is more acceptable to humans if it is somewhat less anthropomorphized unless it is able to so closely approximate human capacities and characteristics as to be indistinguishable.

Likewise, philosophies of religion that admit no anthropomorphism have been unappealing beyond a narrow tract of mystics and intellectuals, for example, Plotinus, Meister Eckhart, and F.D.E. Schleiermacher. By contrast, approaches that either admit modest, qualified amounts of anthropomorphism, for example Augustine or Aquinas, or fully embrace anthropomorphism, for example Karl Barth, have been the most prevalent and most broadly accepted. Meanwhile, it was the Audians in fourth century Syria, whose belief that God has a human form, that were deemed heretics (Patterson 2012), their degree of anthropomorphism falling at the very bottom of the uncanny valley. Contemporary debates in philosophy of religion are similarly so interpretable. Advocates of classical theism denigrate advocates of perfect being theology (PBT) by ascribing the label of 'personal theism', and their criticism is largely that PBT risks falling into, or indeed has already fallen into, the uncanny valley, too closely approximating humanity (Wood 2021). Perfect being theology advocates largely respond by insisting that they are at the pinnacle of acceptability above the cliff edge overlooking the uncanny valley, whereas classical theists are afraid to summit the peak and so remain below the ridgeline of the forest of aseity.

Two of the most trenchant critics of anthropomorphizing divinity in modern philosophy, Hume and Feuerbach, both identified anthropomorphism as the mental projection of human qualities onto the cosmos, and recognized the psychological and social needs fulfilled by religion in doing so (Feuerbach 1857; Hume 2007). Contemporary empirical findings in the cognitive and evolutionary sciences of religion confirm that anthropomorphizing divinity and other supernatural agents is a by-product of evolutionarily adaptive cognitive biases that became culturally adaptive in the context of developing complex human societies (Purzycki and Sosis 2022). Likewise, AI researchers have long recognized the Eliza effect, or the susceptibility of humans to ascribe agency and understanding to even very basic computational systems that produce interpretable text (Hofstadter 1996). Yet an AI engineer at Google became thoroughly convinced that the LLM he had worked on building had in fact become sentient, resulting in his being fired for making unfounded claims (The Guardian 2022). Philosophers of religion would do well to learn from discussions of anthropomorphism in AI, in conversation with the historical debates over anthropomorphizing divinity, and to remember that evolved cognitive biases may lose their adaptive functions as the cultural context shifts, and even become maladaptive. It may be that old

anthropomorphisms are no longer adequate, even if they can be rendered coherent and consistent enough, in which case the field is ripe for renewal. And perhaps we philosophers of religion can help AI researchers avoid our own past pitfalls in the process.

Competing interests. The author declares none.

References

- Feuerbach L (1857) *The Essence of Christianity*, Translated by Evans M. 2nd New York: Calvin Blanchard.
- The Guardian*. 2022. Google Fires Software Engineer Who Claims AI Chatbot Is Sentient, July 23, 2022, sec. Technology. www.theguardian.com/technology/2022/jul/23/google-fires-software-engineer-who-claims-ai-chatbot-is-sentient.
- Hofstadter DR (1996) *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York: Basic Books.
- Hume D (2007). In Coleman D (ed.), *Dialogues Concerning Natural Religion and Other Writings*. Cambridge: Cambridge University Press.
- Leshner J (2023) Xenophanes. Edward NZ, Nodelman U and Summer (eds), *The Stanford Encyclopedia of Philosophy*. 2023 Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2023/entries/xenophanes/>.
- Melanie M and Krakauer DC (2023) The Debate over Understanding in AI's Large Language Models. *Proceedings of the National Academy of Sciences* **120**(13), e2215907120. <https://doi.org/10.1073/pnas.2215907120>
- Mori M (2012) The Uncanny Valley. Translated by MacDorman K and Kageki N *IEEE Robotics and Automation Magazine* **19**(2), 98–100. doi:[10.1109/MRA.2012.2192811](https://doi.org/10.1109/MRA.2012.2192811)
- Patterson PA (2012) *Visions of Christ: The Anthropomorphic Controversy of 399 CE*. Tübingen, Germany: Mohr Siebeck.
- Purzycki BG and Sosis R (2022) *Religion Evolving: Cultural, Cognitive, and Ecological Dynamics*. Sheffield, UK: Equinox Publishing Limited.
- Shensheng W, Lilienfeld SO and Rochat P (2015) The Uncanny Valley: Existence and Explanations. *Review of General Psychology* **19**(4), 393–407. <https://doi.org/10.1037/gpr0000056>.
- Wood W (2021) *Analytic Theology and the Academic Study of Religion*. Oxford: Oxford University Press.