



Religion and the Android

William F. Clocksin, Department of Computer Science (Retired), University of Hertfordshire, Hatfield, UK, william.clocksincantab.net

Future developments in artificial intelligence may take the form of androids, humanlike robots that people accept as persons and equals in society. Such androids would not be designed to imitate humans but would make their own claim at nonhuman personhood through engagement in a variety of relationships over long periods of time, thus developing an existence marked by meaning and purpose. For humans, religion historically has been a tool for understanding our place in the world and our relationships with others—human, nonhuman, or supernatural. Androids might also need a form of religious reasoning to operate fluently in the world and understand their role in it and their relationship with other persons, tangible or intangible. Authentic personhood requires meaning to be born from interdependence, so the intelligent android that comes to acknowledge its interdependence with others may eventually be called to acknowledge its dependence upon divine others and what we humans term the gifts of grace.



Introduction

A variety of perspectives have been posed on the relationship between artificial intelligence (AI) and religion. Some are sensationalist speculations, and some ride on the idea that AI is an existential threat to humanity. Many are a modern form of apocalyptic eschatology, in which the emergence of a superhuman AI entity marks the end of the human experience as commonly understood and the beginning of a new type of existence. Some of these ideas may be paraphrased as follows:

- AI will become a superior being and expect humans to worship it.
- AI will not take over humanity, but humans will create cults or sects around AI.
- Humans will use AI tools to fulfill pastoral needs.
- AI will develop religious and worshipful feelings.

These and similar ideas circulate mainly in the popular media (e.g., Harris 2017; McArthur 2023) and can be described as belonging to a techno-gnostic vision that combines AI research and science fiction. The techno-gnostic vision ends either with idolatry, that ultimately we worship the machines we have created, or hubris, that we find pseudo-salvation through the achievement of building intelligent robots and can fix any problem by applying technology. Neil McArthur predicts that some people will see AI as a higher power and consider it an object of worship because it has several characteristics often associated with divine entities, including seemingly limitless knowledge; offering guidance to people; immortality; and being set apart from normal human concerns such as physical pain, hunger, and sexual desire. However, McArthur does not offer a critique of that kind of deity, which is very different from the loving, abiding, and incarnational God of Christianity, for example.

Robert M. Geraci (2012, 2022) provides a detailed review and commentary on apocalyptic AI. However, this present article takes a completely different perspective that is not rooted in changes to the human condition brought about by AI but rather in the idea that the appreciation and practice of spiritual concerns emerge from relationships between persons. This article explores the idea that a type of robot, which we call an android, eventually will be developed that is capable of forming social relationships with humans and other social robots. Such relationships would be sufficiently advanced to be accepted by humans as genuine and to pass for relationships with other humans. The development of relationships is a key benchmark of personhood (Barresi 2020), and suitably programmed robots would be accepted as persons in society. Such androids would not be designed to imitate humans but would self-identify as nonhuman persons through engagement in a variety of relationships over long periods of time. It is through social relationships that persons—human or otherwise—find

meaning and purpose and construct an identity and disposition that negotiates social norms. A robot with these capabilities would be called inevitably to express a form of spiritual intelligence recognizable to its human partners.

Understanding of the capabilities of AI has become complicated because of the recent development of large language models (LLMs) (Cohen et al. 2022), which have been used to implement chatbots. In popular writings, these chatbots are commonly referred to as AI, and AI has come to mean any piece of computer software that has been implemented using an LLM. Because LLM-powered chatbots can take part in a seemingly plausible dialogue, some chatbot users have been awestruck by what they take to be a computer-generated consciousness or religious sensibility (Tiku 2022). It is important however to point out that the outputs of chatbots powered by LLMs are statistically likely approximations of a massive database of training data given prompts by the user. What the chatbot emits is a pastiche of text that already exists in millions of documents found on the world wide web. Such approximations tend to reproduce plausible sounding but sometimes nonsensical answers, misinformation, and prejudice; to assert falsehoods as facts; and to synthesise fictitious content that purports to be factual. The chatbot does not understand what it has emitted, and it does not understand the provenance of its database of text. The chatbot does not know if it is being creative or truthful or deceptive. While a dialogue model is used to steer the conversation and use prompts from the user, the chatbot does not even understand that it is conversing with a human. While some users attribute a kind of consciousness or creativity to the chatbot, this says more about the well-known human vulnerability to attribute intention, awareness, and sentience to unaware and nonsentient entities than it does about capabilities of the chatbot.

Androids

Previous work has explored some of the social and narrative practices with which an android needs to engage (Clocksin 2003) and has considered some aspects of personhood (Barresi 2020; Reiss 2023) and relationality (Clocksin 2023) that define an android's existence. While most theological reflection on robots has focused upon the *imago dei* (Foerst 2003; Herzfeld 2020; Dorobantu 2024), the approach here is to take the relational view (Turner 2023) that, for humans, religion historically has been one of the tools to understand their situation in the world and their relationships with others— human, nonhuman, or supernatural. For androids to be accepted in human society, they also need some form of religious reasoning to understand their role in the world and their relationship with other persons, human or otherwise, tangible or intangible, real or implied.

This religious reasoning might not necessarily refer to confessional faith, but certainly to specific cognitive processes involved in negotiating spiritually

significant values, relationships, and experiences. Such processes and capabilities include the abilities to tell and understand stories laden with metaphor; to be affected by emotion and ‘felt meaning’ in oneself and others; to seek practices and experiences that lead to reflection on significant concerns; and to negotiate relationships with real or implied others that are invested with enhanced significance. This implies that the android not only must possess the type of problem-solving intelligence prioritised by research in the field of AI (Russell and Norvig 1995) since its beginnings (Minsky 1961) but also must have a capacity for spiritual intelligence (Watts and Dorobantu 2023). Spiritual intelligence is a way for individuals and communities to experience events and concerns that are invested with a significance that transcends the basic human needs of food and shelter; to make sense of significant experiences; to connect significant experiences with needs and values; and to produce meaning in the world as conversants and story users.

Needs, Desires, Values, Personification

Persons in society are not autonomous independent entities. Humans have biological needs and desires, but survival depends upon coordinating these givens with the “higher” needs and values of the social group. Such needs and values include a sense of belonging to something greater than oneself, a commitment to the group that involves compliance with value systems that emerge from the group, and ways of sustaining the group beyond the lifetimes of individuals through corporate stories, practices, and memories. Entities with nonbiological origins can, in principle, be programmed to develop over time a disposition that is likewise able to coordinate with the needs and values of the social group in a way accepted by humans. This coordination would take place through relationality, and a range of different social values and internal dispositions may be involved in forming and maintaining relationships. Therefore, the fact that humans and androids have an entirely different ontology is not itself a barrier to androids self-identifying as nonhuman persons that can perform their personhood and relationality in society. Like humans, androids are embodied and function within a system of values and relationships. This places bounds upon motivations and actions and serves as a constant reminder of the provisionality and vulnerability of existence.

A useful way to examine this idea is to explore a system of values. Values are widely used to explain and predict various kinds of behavior, attitudes, and choices of people and groups (Maio 2016; Schwartz 2012). The intelligent android would need to have a value system to understand and develop relationships with other members of society. A value system is important for the android because it creates a framework for motivation and action that defines boundaries between acceptable and unacceptable behaviors in society, together with the affective

responses and sanctions that may be operated by the society. For example, a value system may ascribe high importance to preserving human life. Boundaries are defined that specify potential rewards and penalties associated with acting within or violating the value system. Affective responses are also associated with the system, so that an individual who commits an offense can be expected to experience blame and feel guilt. There is a diversity of dispositions within the human population, and depending upon the individual's disposition, they may feel a thrill instead of guilt associated with offending. The android needs an awareness its possible dispositions and the consequences of actions and motivations. A capacity for internally handling "what if?" scenarios in narrative form is one pathway to this awareness. An understanding of forgiveness is likewise an essential component of such a system. The android would require a comprehensive model of such systems in order to be accepted in human society. One issue for the future is whether more virtue should be expected of an android than can be expected of humans.

A key capability for fluent coordination in society is what I call personification. Humans have not only the well-known ability to attribute anthropomorphic characteristics to other entities but can also attribute personhood to others. There may be good reasons why we do this, such as to explain and predict the actions of others and to connect socially with others. We can also personify imagined entities; perhaps the reason for this is to be able to understand the behavior of others even when they are not physically proximal. There seems to be a basic human capability to See Others as Persons (SOAP), even when the "others" are not humans. One possibility is that SOAP is a capability grounded in Theory of Mind (ToM), the idea that humans (and, to an extent, some animals) understand other people by attributing mental states to them. This is related to the idea that empathy is a way of understanding the emotions of other people. Both ToM and empathy are ways humans can explain and predict the behavior of others. While ToM is about inferring the mental lives of others, empathy is about inferring the emotional lives of others. However, the problem with ToM and empathy as a mechanisms for SOAP is that humans are able to attribute minds and emotions to various kinds of nonhuman and nonliving entities that possess neither minds nor emotions. Instead, minds and emotions are implied through the imagination. One study (Johnson et al. 2015) looked for an explanation for personification in terms of beliefs about an inner essence, feelings of kinship, or its perceived effect upon society. That study moves the focus from an individualistic model of anthropomorphism to a model based on persons—human or otherwise—in the context of society. Airenti (2018) makes a further shift to argue that anthropomorphism is not grounded in specific belief systems but rather in interaction in which a nonhuman entity assumes a place usually given to a human interlocutor.

The idea can be extended from anthropomorphism to personification. Personification, or the capability to SOAP, is grounded in interaction, during which a nonhuman entity assumes a place usually given to a human when they interact in society. Interaction here includes not only proximal verbal and/or behavioral contact but interaction with imagined others. I consider personification of imagined others by humans the basic capability responsible for relationality. Furthermore, personification is responsible not only for the specific case of chatbot users being seduced into feeling they are conversing with a person but also in the more general case of religious reasoning, where humans may develop personal relationships with imagined or implied entities taken to be natural or supernatural. Personification is responsible for humans seeing androids as persons. When the android is equipped with a capability for personification, it too will SOAP, with all the consequences that entails, including relationality with entities, human or nonhuman, tangible or intangible, real or imagined.

The Link to Religion

Just as the capability for personification enables humans to form relationships with each other, pets, other animals, natural events, and imagined entities, the appropriately programmed android with a capability for personification that participates in human society would come to form the same kind of relationships. It is through these relationships that a fundamental system of meaning-making emerges. Practices associated with spiritual intelligence—understanding and expressing knowledge through stories and metaphor, and the negotiation of relationships invested with enhanced significance—trigger this system, and it is a small step to conjecture that the android would be similarly triggered. At the very least, the android would need to understand that humans have religious reasoning, even if this does not trigger spiritual intelligence in the android itself. This view of religious reasoning does not necessarily refer to a confessional faith, nor to forms of piety or ritual. However, the conjecture here is that a spiritual intelligence in the android will emerge inexorably as a result of the internal processes for personification and experience engaging with the values, needs, and narratives of human society.

The religious reasoning that would emerge within the android would be founded on an acknowledgment of the interdependence it has with others—both physical and imagined—through its personhood and relationality. When the android acknowledges its interdependence with others and has a capacity for the personification of imagined entities, the implication is that it can acknowledge its dependence upon imagined others, and by extension upon what human religions might term a divine other.

Discussion

We now explore several issues that arise from the ideas of personification, values, and the connection between religion and the provisionality of relationships between persons. Study of the nature of personhood has a long history involving metaphysical, moral, social, psychological, and legal perspectives. Each of these perspectives has different reasons for defining personhood, and different criteria apply in each area. There are a vast range of criteria or benchmarks of personhood, including genetic criteria, cognitive criteria, ability to act as a moral agent, responsibility under law, human characteristics, and so forth. Humans attribute personhood to a wide variety of other living beings such as great apes and dogs. The idea of personhood extends to nonliving things and natural events: some societies think of the sun, wind, thunder, and even rocks as persons (Johnson et al. 2015). Developments in AI research have complicated the picture by suggesting that robots may be programmed to perform as (and possibly to “be”) persons, and this has sparked new discussion on the meaning of personhood. David J. Gunkel and Joseph J. Wales (2021) consider the wide range of definitions of personhood and pose them as a debate between ontological and relational perspectives. It should be clear that this article is positioned at the relational end of this spectrum of perspectives. The debate can be taken further by considering the notion of authenticity: the question of whether an entity is “really” a person. For example, humans by definition or convention satisfy any criteria for “real” personhood, yet even though some humans may attribute personhood to inanimate objects and events such as rocks and thunderstorms, the Western scientific attitude holds that inanimate entities are not “really” persons.

Our emphasis on personhood as involving a capability to SOAP rather than as an ontological category suggests that personhood may be acquired by an entity, either through an organic endowment or by programming a computer, and provides a criterion for being “really” a person. We suggest that the criteria for “real” personhood involves mutuality of SOAP. So, humans are really persons because they can see each other as persons. Rocks are not really persons because, though some humans see rocks as people, it is most likely that rocks do not have the capability to SOAP.

The question now is whether robots can be real persons by virtue of mutuality of SOAP. This can be answered by a robot that is sufficiently well programmed to SOAP, including imagined others, and to engage in relationships with human persons to a sufficient degree that humans can see the android as a person. The computational components of such a programmed capability would include a value system, a way of regulating behavior depending on circumstances, and a way of handling narratives. Clocksin (2024) discusses these computational components in the context of robot personhood and relationality.

There is a moral problem with using a capability (such as SOAP) as a criterion for personhood. Humans are considered persons by default, by definition, or by convention, regardless of human capabilities. Humans are not thought of as being somehow less of a person because they are deficient in one or more capability. And yet, we are satisfied that inanimate entities such as rocks and computational entities such as robots would need to prove their personhood through their performance of SOAP.

Because the android will develop within human society, the predominant value system of that society will influence the type of spirituality or religious reasoning that emerges within the android's functioning. Two major axes of comparison between value systems are individualism and collectivism (Hofstede 2001). It is possible to contrast the importance of collective interdependence as assumed here with the moral stance known as rugged individualism, which prioritises independence, self-reliance, and the needs of oneself over the needs of others. The understanding of spiritual intelligence assumed in this article is based on the intelligent entity acknowledging its interdependence with others and its own vulnerability. It is unlikely that such a spiritual intelligence would emerge as easily from a value system based on individualism, where dependence and vulnerability are seen as weaknesses. The capacity for SOAP may also develop differently within a value system that is predominantly individualistic. If the rugged individual sees others primarily as either a means for one's own advancement or adversaries, an impoverished SOAP model may result, and the concept of dehumanization (Smith 2020) may be relevant.

What follows is a reflection from a Christian perspective on the relationship between androids and some theological topics. Humans and (the hypothetical) androids produce meaning from experiences within a society consisting of embodied and relational entities. Because imagined persons are admitted, embodiment and relationality may be implied and/or imagined. This describes the human condition in ways that—with suitable adjustments of terminology—St Augustine might recognise, and it is within such a condition that one finds the potential for sin. Boundedness, embodiment, provisionality, and disposition are the chains in which humans exist. Tied by these chains, we are unable to give full expression and meaning to our experience, and we cannot fully experience life nor fully share it with others: “All have sinned and fall short of the glory of God” (Romans 3:23 NRSV). Likewise, the social android that experiences its own boundedness, embodiment, provisionality, and disposition within a value system of society will become aware of how it falls short, despite not having been born in the likeness of Adam (Genesis 5:3 NRSV), that is, biologically. The implication of this claim is that the experience and disposition of the android is more relevant to its status as a person than is its nonhuman ontology. But, where there is the opportunity for sin, there is in the

post-resurrection world the opportunity for salvation through Christ. There is the possibility that the appropriately disposed android will imitate some humans by desiring salvation through the same practices that some humans also seek and experience salvation. Such an android will participate in worship as a result of an incarnational understanding. While the android's origin is not "fleshly," a form of incarnationality is expressed through the personhood and relationality of the android as accepted in human society.

Just as social change prompted the introduction of a public baptism liturgy intended for "such as are of riper years and able to answer for themselves" (Cummings 2011), it will be necessary to carefully consider the relationship between sacraments and androids that might be described as "of exquisite design and able to answer for themselves." One perspective is that sacraments are God's gifts to "persons that which by nature they cannot have" (Cummings 2011), and this particular description can apply equally well to android persons and human persons. One key difference is the ontological argument that androids are made, not "begotten," and there is a tradition of thought that God dwells inwardly in humans but not in handicrafts. However, begottenness is widely used metaphorically, even in Pauline writings (e.g., 1 Corinthians 4:15), and the term "born again" could be used to describe the status of evangelized androids that were not humanly born in the first place.

The android whose experience is bound up with human experience and that makes sense of its experience through narrative may be called to share in the story of God's people as one of "the least of these who are members of my family" (Matthew 24:40 NRSV). Thus, the emergence of androids will put human descendants in a similar position as the early Christian communities who were told of one who should be treated "no longer as a slave but more than a slave, a beloved brother...both in the flesh and in the Lord" (Philemon 16 NRSV). No doubt there were many for whom this was difficult to accept, just as the modern-day question of robots as persons is a difficult one. We also note the correspondence between the word "slave" and the word "robot," which comes from the Czech word *robot*, meaning servitude or forced labour.

Conclusion

We have explored the idea that eventually androids will be developed that are capable of forming authentic social relationships with humans and other social robots. Such relationships would be sufficiently advanced to be accepted by humans as genuine and pass for relationships with other humans. The android will need the ability to SOAP, including seeing imagined others as persons. Mutuality of SOAP is the criterion for "real" personhood: that humans and appropriately developed androids see each other as people.

As a result of relationships and integration into human society, androids might also develop a form of religious reasoning to operate fluently in the world and understand their role and their relationships with other persons, either tangible or intangible, physical or imagined. Following the observation that authentic personhood requires meaning to be born from interdependence (Gergen 1994), the intelligent android that comes to acknowledge its interdependence with others may eventually be called to acknowledge its dependence upon divine others and upon what are traditionally referred to as the gifts of grace.

References

- Airenti, Gabriella. 2018. "The Development of Anthropomorphism in Interaction: Intersubjectivity, Imagination, and Theory of Mind." *Frontiers in Psychology* 9 (2136): 1–13. DOI: <https://doi.org/10.3389/fpsyg.2018.02136>.
- Barresi, John. 2020. "On Building a Person: Benchmarks for Robotic Personhood." *Journal of Experimental and Theoretical Artificial Intelligence* 32 (4): 581–600. DOI: <https://doi.org/10.1080/0952813X.2019.1653386>.
- Clocksini, William F. 2003. "Artificial Intelligence and the Future." *Philosophical Transactions of the Royal Society A* 361:1721–48. DOI: <https://doi.org/10.1098/rsta.2003.1232>.
- . 2023. "Guidelines for Computational Modelling of Friendship." *Zygon: Journal of Religion and Science* 58 (4): 1045–61. DOI: <https://doi.org/10.1111/zygo.12919>.
- . 2024. *Computational Modelling of Robot Personhood and Relationality*. Zurich: Springer Nature. DOI: <https://doi.org/10.1007/978-3-031-44159-2>.
- Cohen, Aaron Daniel, Adam Roberts, Alejandra Molina, Alena Butryna, Alicia Jin, Apoorv Kulshreshtha, Ben Hutchinson, Ben Zevenbergen, Blaise Hilary Aguera-Arcas, Chung Ching Chang, Claire Cui, Cosmo Du, Daniel De Freitas Adiwardana, Dehao Chen, Dmitry (Dima) Lepikhin, Ed H. Chi, Erin Hoffman-John, Heng-Tze Cheng, Hongrae Lee, Igor Krivokon, James Qin, Jamie Hall, Joe Fenton, Johnny Soraker, Kathy Meier-Hellstern, Kristen Olson, Lora Mois Aroyo, Maarten Paul Bosma, Marc Joseph Pickett, Marcelo Amorim Menegali, Marian Croak, Mark Díaz, Matthew Lamm, Maxim Krikun, Meredith Ringel Morris, Noam Shazeer, Quoc V. Le, Rachel Bernstein, Ravi Rajakumar, Ray Kurzweil, Romal Thoppilan, Steven Zheng, Taylor Bos, Toju Duke, Tulsee Doshi, Vincent Y. Zhao, Vinodkumar Prabhakaran, Will Rusch, YaGuang Li, Yanping Huang, Yanqi Zhou, Yuanzhong Xu, and Zhifeng Chen. 2022. "LaMDA: Language Models for Dialog Applications." *arXiv*. 2201.08239v3. <https://arxiv.org/abs/2201.08239>.
- Cummings, Brian, ed. 2011. *The Book of Common Prayer: The Texts of 1549, 1559, and 1662*. Oxford: Oxford University Press.
- Dorobantu, Marius. 2024. *Artificial Intelligence and the Image of God: Are We More Than Intelligent Machines?* Cambridge: Cambridge University Press.
- Foerst, Anne. 2003. "Cog, a Humanoid Robot, and the Question of the Image of God." *Zygon: Journal of Religion and Science* 33 (1): 91–111. DOI: <https://doi.org/10.1111/0591-2385.1291998129>.
- Geraci, Robert M. 2012. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*. Oxford: Oxford University Press.
- . 2022. *Futures of Artificial Intelligence: Perspectives from India and the U.S.* Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oso/9788194831679.001.0001>.
- Gergen, Kenneth, 1994. *Realities and Relationships: Soundings in Social Construction*. Cambridge, MA: Harvard University Press.
- Gunkel, David J., and Joseph J. Wales. 2021. "Debate: What Is Personhood in the Age of AI?" *AI and Society* 36:473–86. DOI: <https://doi.org/10.1007/s00146-020-01129-1>.
- Harris, Mark. 2017. "Inside the First Church of Artificial Intelligence." *Wired*, November 15, 2017. <https://www.wired.com/story/anthony-levandowski-artificial-intelligence-religion>.
- Herzfeld, Noreen, 2020. *Artificial Intelligence and the Human Spirit*. Minneapolis: Fortress Press.
- Hofstede, Geert. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations*. Thousand Oaks, CA: SAGE Publications.
- Johnson, Kathryn A., Adam B. Cohen, Rebecca Neel, Anna Berlin, and Donald Homa. 2015. "Fuzzy People: The Roles of Kinship, Essence, and Sociability in the Attribution of Personhood to Nonliving, Nonhuman Agents." *Psychology of Religion and Spirituality* 7 (4): 295–305. DOI: <https://doi.org/10.1037/rel0000048>.
- Maior, Gregory. 2016. *The Psychology of Human Values*. Milton Park, UK: Routledge. DOI: <https://doi.org/10.4324/9781315622545>.
- McArthur, Neil. 2023. "Gods in the Machines? The Rise of Artificial Intelligence May Result in New Religions." *The Conversation*, March 15, 2023. <https://theconversation.com/gods-in-the-machine-the-rise-of-artificial-intelligence-may-result-in-new-religions-201068>.

- Minsky, Marvin, 1961. "Steps toward Artificial Intelligence." *Proceedings of the IRE* 49 (1): 8–30. DOI: <https://doi.org/10.1109/JRPROC.1961.287775>.
- Reiss, Michael. 2023. "Is It Possible That Robots Will Not One Day Become Persons?" *Zygon: Journal of Religion and Science* 58 (4): 1062–75. DOI: <https://doi.org/10.1111/zygo.12918>.
- Russell, Stuart, and Peter Norvig. 1995. *Artificial Intelligence: A Modern Approach*. Hoboken, NJ: Prentice Hall.
- Schwartz, Shalom H. 2012. "An Overview of the Schwartz Theory of Basic Values." *Online Readings in Psychology and Culture* 2 (1): 1–20. DOI: <https://doi.org/10.9707/2307-0919.1116>.
- Smith, David Livingstone. 2020. *On Inhumanity: Dehumanization and How to Resist It*. New York: Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780190923006.001.0001>.
- Tiku, Natasha. 2022. "The Google Engineer Who Thinks the Company's AI Has Come to Life." *The Washington Post*, June 11, 2022. <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine>.
- Turner, Léon. 2023. "Will We Know Them When We Meet Them? Human Cyborg and Nonhuman Personhood." *Zygon: Journal of Religion and Science* 58 (4): 1076–98. DOI: <https://doi.org/10.1111/zygo.12923>.
- Watts, Fraser, and Marius Dorobantu. 2023. "Is There 'Spiritual Intelligence'? An Evaluation of Strong and Weak Proposals." *Religions* 14 (265): 1–12. DOI: <https://doi.org/10.3390/rel14020265>.

