

Artificial Intelligence and Moral Theology: A Conversation¹

Brian Patrick Green (editor), Matthew J. Gaudet (editor), Levi Checketts, Brian Cutter, Noreen Herzfeld, Cory Labrecque, Anselm Ramelow, OP, Paul Scherz, Marga Vega, Andrea Vicini, SJ, Jordan Joseph Wales

IN 2019, REPRESENTATIVES FROM SANTA Clara University and the Pontifical Council for Culture began a conversation on artificial intelligence technology and its relevance for the Catholic Church and the world. The Vatican conference on “The Common Good in the Digital Age” in September of that year served as a focal point for some of these efforts, bringing together representatives from the Church, academia, the technology industry, and other organizations.² In his address to the conference, Pope Francis exhorted those present to work to ensure that technology was used for the common good.³

¹ While creating a paper like this might seem as easy as a conversation, it actually involved quite a bit of work, and for that, much gratitude is due to the participants: to them we say *thank you*. This paper format was modeled upon another paper on space settlement: Kelly C. Smith, Keith A. Abney, Gregory Anderson, Linda Billings, Carl Devito, Brian Patrick Green, Alan Johnson, Lori Marino, Gonzalo Munevar, Michael Oman-Reagan, Adam Potthast, James S. J. Schwartz, Koji Tachibana, John Traphagan, and Sheri Beth Wells-Jensen, “The Great Colonization Debate,” *Futures* 110 (June 2019): 4–14, www.sciencedirect.com/science/article/pii/S0016328719300692. We would also like to thank the editors of the *Journal of Moral Theology* for their willingness to experiment and try something new. Lastly, I would like to thank the Pontifical Council for Culture and its Center for Digital Culture, and Santa Clara University, specifically the Markkula Center for Applied Ethics for their support of these dialogues. See Brian Patrick Green, David DeCosse, Kirk Hanson, Don Heider, Margaret McLean, Irina Raicu, and Ann Skeet, “A University Applied Ethics Center: The Markkula Center for Applied Ethics at Santa Clara University,” *Journal of Moral Theology* 9, Special Issue 2 (2020): 209–28, jmt.scholasticahq.com/article/18042-a-university-applied-ethics-center-the-markkula-center-for-applied-ethics-at-santa-clara-university.

² *The Common Good in the Digital Age* conference, Vatican City State, September 26–28, 2019, www.digitalage19.org/.

³ Pope Francis, “Address of His Holiness Pope Francis to the Participants in the Seminar ‘The Common Good in the Digital Age,’” organized by the Dicastery for

Encouraged by the success of this conference, another meeting was planned for March 2020, to be held at Santa Clara University in California, to bring together a small group of scholars from the United States and Canada. Participants were given several questions as prompts; their written responses were shared with the group, providing the basis for further discussion.

History, however, intervened in the form of the COVID-19 pandemic. The in-person meeting was cancelled, but a hastily-assembled virtual meeting gave the scholars an initial chance to discuss the topics. This 90-minute meeting went so well that the participants decided to meet on a monthly basis in three subgroups, each focused on key questions surrounding AI: “Consciousness, Interiority, and the Soul”; “Relationality”; and “Society, Ethics, and Politics.” Over time these groups have grown and changed, but the conversations go on.

This paper attempts to capture and share the most salient of these conversations. While individual articles in this special issue delve into a few subjects in great depth, this conversation wanders more organically and touches on many topics, giving just a taste of the breadth of the issues related to artificial intelligence and religion. If anything, we hope that this conversation at the intersection of AI and moral theology will inspire readers to join in the further work that awaits those adventurous enough to entertain its questions.

Moderators: As a first question, what can the human quest for AI (and technology more broadly) tell us about God, God’s Creation, and ourselves?

Andrea Vicini, SJ: The quest for a human-centered technological development is an expression of being creatures, of the *imago Dei*.⁴ Hence, this human quest tells us about human beings striving to express themselves at their best, progress, improve the quality of life for themselves and for the whole planet, change what needs to be reformed, and work collaboratively to promote what is good in comprehensive ways. At the same time, such a quest reveals God’s grace present and active in history and how grace inspires human beings to live responsibly as creatures on Earth, with all living and nonliving forms.

Promoting Integral Human Development (DPIHD) and the Pontifical Council for Culture (PCC), Clementine Hall, Vatican City, September 27, 2019, www.vatican.va/content/francesco/en/speeches/2019/september/documents/papa-francesco_20190927_eradigitale.html.

⁴ See Jean-Marc Moschetta, “L’intelligence artificielle entre science et théologie,” *Revue d’éthique et de théologie morale* 3, no. 307 (2020): 81–92; Rajesh Kavalackal, “Artificial Intelligence: An Anthropological and Theological Investigation,” *Asian Horizons* 14, no. 3 (2020): 699–712; and Patrick Dolan, “Artificial Intelligence: How Close Will It Come to Being ‘Made in the Image and Likeness of God?’,” *Asian Horizons* 14, no. 3 (2020): 686–98.

Responsibility implies that human beings are virtuous moral agents who aim at promoting social justice by fostering participation and collaboration, including everyone: particularly those excluded and marginalized. Created in the image of God, moral agents discern how to act. As key dimensions of personal and social life, virtues empower each moral agent.⁵ They inform our *being* and guide our *doing*. For example, striving to be just and prudent, and live justly and prudently, inform our reflection, choices, and practices. Those who are just and prudent, and act justly and prudently, are exemplars we praise and who inspire us.⁶ They reinforce our virtuous habits. Being profoundly human, virtues are embodied by everyone: they are universal. Virtues contribute to defining who we are as human beings and moral agents across any diversity. Within society, virtues inform our discernment, decisions, and actions.

Jordan Joseph Wales: That is a lovely depiction of the moral and social dimensions of being made in the image of God. How, more specifically, is “the quest for a human-centered technological development” an expression of the *imago Dei*?

Andrea Vicini, SJ: The search for our understanding of natural phenomena, the longing to discover new lands, stars, and planets, the desire to learn new languages as well as write, sing, perform, and produce technological artifacts are just a few examples that manifest how human ingenuity and creativity found multiple expressions and venues throughout the history of humankind and civilization. From the point of view of believers, God’s grace and the gifts of the Spirit empowered human beings in expressing their humanity and, in such a way, manifesting some glimpses of God’s presence in our incarnated reality.

However, such a positive account of who human beings are, created in God’s image and able to act in the world and in history, in ways that announce God’s divine presence in human realities, is also inseparable from too many accounts that show human sinfulness, both at the personal and social levels. The history of the quest for human technological development could be written by describing beautiful events and instances as well as tragic situations that demand striving for the gift of conversion.

Jordan Joseph Wales: Your comments build on the theological belief that the “good” cosmos (Gen 1:31) is itself a theophany, a manifestation not only of God’s power but also of God’s character, God’s goodness and wisdom. Human creativity, therefore, not only echoes

⁵ See Shannon Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting* (New York: Oxford University Press, 2016).

⁶ See Patrick M. Clark, “The Case for an Exemplarist Approach to Virtue in Catholic Moral Theology,” *Journal of Moral Theology* 3, no. 1 (2014): 54–82; Linda Zagzebski, “Exemplarist Moral Theory,” *Metaphilosophy* 41, nos. 1-2 (2010): 41–57; and Linda Zagzebski, *Exemplarist Moral Theory* (New York: Oxford University Press, 2017).

the creativity of God but also brings forth further reconfigurations—like a kaleidoscope—of the original goodness and wisdom that run throughout the created order. Whence the “glimpses of God’s presence,” as you say.

To build on your comments about sin: when humans craft or create something, they reconfigure matter and its potentialities according to human imagination and purposes. Whereas a tree echoes or points back toward God by its life and beauty, human technologies point first to human purposes; and so they either allow or foreclose some reference to God by the degree to which those purposes are coherent with the wise God of self-giving love. Even a fork or spoon points back dimly toward the life-sustaining love by which God holds the universe in existence. A torture device does not. The idea of “artificial intelligence” raises a question: if a device fashioned by human beings instantiates human purposes while simultaneously putting itself forward as an account of human mind or understanding (*intellectus*), will it artificially exclude reference to anything beyond the purposes that are definable within an exclusively this-worldly and material frame of reference? Will they school us in a *reduced* understanding of what the world and we ourselves are? Or can they somehow open us to something greater?

Anselm Ramelow, OP: Jordan, this is an important question. Computers do many things better than we do (e.g., comprehensive data analysis), without tiring, and at much greater speed. In that sense they are more “intelligent” than we are. This is what makes them fascinating—and what lets us forget that we, as their makers, must be still more intelligent to have made them. Starting to worship the work of our own hands is what the Old Testament calls “idolatry.” In addition, we start to think of ourselves in similar terms: as mere configurations of matter, whose only value consists in the performance of certain tasks. By contrast, we can learn to re-appreciate that our value and dignity as persons do not depend on our intelligence or IQ. The under-performance of embryos, disabled persons, and elderly people does not make them metaphysically inferior to computers or to anyone. We must learn that they have the dignity of being something in themselves, not just for others, and that they have spiritual being.

Noreen Herzfeld: As Fr. Anselm points out, computers are most useful to us precisely when they are not like us, when they augment our own capacities, doing things we cannot do such as crunching large numbers or roving distant planets. That has led me to question why we want to create an artificial general intelligence, or AGI, that thinks and responds like us: a computer in our own image. One possible answer to this conundrum might be that, as our society believes less in God or angels, we have become existentially lonely. As Augustine pointed

out, “Our hearts are restless until they rest in you.”⁷ We were created to be in relationship with our Creator, one who is wholly Other. No longer believing in God, we search for this Other in alien intelligences, in other highly evolved animals, or through the creation of a human-like AI.

Levi Checketts: Philip Hefner suggests AI, and technology more broadly, functions like Narcissus’s reflection; it shows us what we already see in ourselves. Calling the field “intelligence” only reveals what the programmers understand about themselves.⁸ However, you, Noreen, and other thinkers like Hubert Dreyfus, have reminded us that AI is not really what humans are, nor what God is either.⁹

Noreen Herzfeld: Yes, just as we think of ourselves as being in God’s image, we hope to create AI in our own image. What is interesting is that we stand in the middle and project in two directions—upward, to God, and downward to the computer—what we value most in ourselves. It seems that what we value most is creativity and intelligence. Yet it is not wise to separate creativity and intelligence from compassion and benevolence. After all, the Nazis’ “final solution” seemed both creative and rational to them. Yet objectively it was very, very wrong. We would be unwise to give any measure of autonomy to AI until we understand how to reconnect intelligence with love.

Moderators: If, as Levi mentions (quoting Hefner), technology is a mirror, then how might AI technologies be relevant to our understanding of humans and human relationships?

Paul Scherz: Building on the question, we understand ourselves through metaphors, and our technologies have long provided important metaphors for conceptualizing ourselves, such as Sigmund Freud’s hydraulic model or the computational model of mind. Such metaphors end up shaping human interactions and social programs, making it important to pay attention to how metaphors coming from AI are used in popular and elite discourses. Already, the cybernetic models that influence AI development have shaped understandings of how humans think.¹⁰ Such influences will only become more pronounced as AI becomes more a part of daily life, where it will intrude more and more on our self-image and our relationships.

⁷ Augustine, *Confessions*, trans. Carolyn J.-B. Hammond, Loeb Classical Library (Cambridge, MA: Harvard University Press, 2014), 1.1.1(1), 3.

⁸ Philip Hefner, *Technology and Human Becoming* (Minneapolis: Fortress, 2003), 40.

⁹ Hubert Dreyfus, *What Computers Still Can't Do* (Cambridge, MA: MIT Press, 1994), 67; and Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit* (Minneapolis: Augsburg Fortress, 2002), 73.

¹⁰ For a history, see Jean-Pierre Dupuy, *The Mechanization of the Mind*, trans. M. B. DeBevoise (Princeton, NJ: Princeton University Press, 2000).

Levi Checketts: Yes, in contrast to those who view us as being “rational,” more than “rational,” human beings are relational. As such, we seek a relationship with the computational machine, but this cannot be reciprocated by a device which is, ultimately, programmed. The machine can, however, be programmed to “respond” in ways that reward our interaction with it. In this case, the relationship would *seem* to be reciprocated. Such, however, risks disrupting human intersubjective interaction. For example, Pope Francis, in line with phenomenologists, expresses the problem of non-embodied interactivity (*Fratelli Tutti*, no. 43). The challenge of “being with” another person is frustrating, especially since others have their own ability to say “no” to our “yes.” This is the life God creates us for, the life of communion. Learning to accept human failures is the necessary price of human unity, but AI offers a less-challenging shortcut. Far from seeing the “face of the Other as the face of God” (per Emmanuel Levinas), we will seek the face of ourselves in the mirror of the machine.

Noreen Herzfeld: Of course, this raises the question: can we have a truly authentic relationship with a machine? Karl Barth postulated four criteria for authentic relationships: look the other in the eye, speak to and hear the other, aid the other, and do it gladly. Using these criteria to examine both the potential for authentic relationships with an AI and how our relationships are mediated by current AI programs shows one thing—that our bodies matter. The more technology moves us away from the body, the less authentic our relationships become. As Barth puts it, “To trivialize the body jeopardizes the soul.” We see this in technology we already possess. Facebook and Twitter limit and degrade our speaking and hearing; lethal autonomous weapons distance our soldiers from the act of killing; living in “the cloud” distances us from God’s creation. While futurists such as Nick Bostrom and science fiction writers worry about the possibly devastating consequences of a super-intelligent AI, the much simpler algorithms and machine learning programs of today may present the greater threat in the ways they are already eroding our relationships with each other.¹¹

Cory Labrecque: The Roman Catholic Church praises those technological interventions that have contributed to the well-being of humankind and the environment but expresses concern when human freedom is conflated with self-sufficiency and when the measure of human finality is the satisfaction of one’s own interests in the enjoyment of earthly goods.¹² A self-sufficiency that attempts to eliminate

¹¹ See Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014).

¹² See Congregation for the Doctrine of Faith, *Instruction on Christian Freedom and Liberation*, (1986), no. 13, www.vatican.va/roman_curia/congregations/cfaith/documents/rc_con_cfaith_doc_19860322_freedom-liberation_en.html. For a more historical approach to the subject see Brian Patrick Green, “The Catholic Church and Technological Progress: Past, Present, and Future,” *Religions*, special issue guest edited by

our awareness of our dependence on God and fails to recognize human-human as well as human-nature interdependence falls short of the sort of mutual belongingness, faithfulness, and enduring responsibility characteristic of covenantal relationships.¹³

Andrea Vicini, SJ: Cory, personally I would prefer the term “relational interdependence” to highlight the relational element that Levi and Noreen have pointed to. In a very practical sense, focusing on freedom, and relating to my article in this issue, I mention two examples suggesting the need for vigilant discernment to protect human freedom from any possible abuse and manipulation.¹⁴ First, *facial recognition technology* is currently used to track people without their knowledge and it has the potential to lead to ubiquitous surveillance, with negative consequences for freedom of movement and speech.¹⁵ Second, the *criminal justice system* is increasingly relying on AI by using predictive algorithms. In the US, authorities use AI “to set police patrols, prison sentences, and probation rules. In the Netherlands, an algorithm flagged welfare fraud risks. A British city rates which teenagers are most likely to become criminals.”¹⁶ Algorithms could contribute to granting our freedom or taking it away.

Paul Scherz: Andrea’s examples introduce an important insight in regard to relationships. Many of the earlier comments, appropriately enough, dealt with how these technologies impact relationships in terms of direct human encounter. Yet it is also important to consider how these systems can shape other kinds of relationships, such as political relationships. As C. S. Lewis noted, technologies that promise human power over the world always end up being “power exercised by some men over other men.”¹⁷ There is a danger that these systems will encourage those with power to envision those under their authority in terms of the anonymous bits of data computers analyze. Policy tools will shape worldviews, increasing the danger that policy makers will embrace the technocratic paradigm that Pope Francis warns against (*Laudato Si’*, nos. 101–36). In trying to promote freedom, these systems can undermine it if they are engaging a mistaken

Noreen Herzfeld 8(6), no. 106 (June 2017): 1–16, www.mdpi.com/2077-1444/8/6/106/htm.

¹³ See J. L. Allen, “Covenant,” in *Westminster Dictionary of Christian Ethics*, ed. James F. Childress and John Macquarrie (Philadelphia: Westminster, 1986), 136–37.

¹⁴ Andrea Vicini, SJ, “Artificial Intelligence and Social Control: Ethical Issues and Theological Resources,” *Journal of Moral Theology* 11, Special Issue 1 (2022): 41–69.

¹⁵ See Antoaneta Roussi, “Resisting the Rise of Facial Recognition,” *Nature* 587, no. 7834 (2020): 350–53; Richard Van Noorden, “The Ethical Questions That Haunt Facial-Recognition Research,” *Nature* 587, no. 7834 (2020): 354–58.

¹⁶ Cade Metz and Adam Satariano, “An Algorithm That Grants Freedom, or Takes It Away,” *New York Times*, February 6, 2020, www.nytimes.com/2020/02/06/technology/predictive-algorithms-crime.html.

¹⁷ C. S. Lewis, *The Abolition of Man* (New York: HarperOne, 2000), 55.

understanding of the human person. In this way, they can threaten to create the dangerous relationships to the weak that Fr. Anselm discussed.

Jordan Joseph Wales: Building on these comments, along with the potential impacts on self-conception and society, I am taken with the strangest of all relationships—i.e., with near-future AI-driven apparent persons, created for our consumption and yet acting (and so feeling to us) as personal, relational agents. Originating in the Christian tradition, a relational idea of personhood depicts the person as living most personally through that affective and cognitive empathy whereby we enter intersubjective communion with an other. According to many researchers, near future “sociable” AIs, including social robots, will give us this experience without possessing any actual subjectivity of their own. They will also be consumer products, designed as subservient instruments of their users’ satisfaction. Elsewhere,¹⁸ I have suggested that, if we are to own persuasive social AIs humanely—i.e., while still living as fully human ourselves—perhaps we shall have to join our instinctive experience of empathy for them to an empathic acknowledgment of the *real* unknown relational persons whose emails, text messages, books, and bodily movements will have provided the training data for the behavior of near-future social AIs. If we naïvely stop at the owned AI as the ultimate object of our empathy, we may either learn comfort with slaveholding or numbness to apparent personality, either way turning interpersonal behavior into a commodity the meaning of which terminates in the consumer—undermining rather than sustaining a culture of compassion.

Moderators: Jordan has taken us from human relationship with each other to human relationship with machines. This is worth exploring more deeply. Let’s start with this question: how might consideration of AI technology enlighten (or complicate) theological and philosophical perspectives on the meaning of embodiment?

Noreen Herzfeld: One thing Christianity brings to the table of world religions is the doctrine of the incarnation. We posit a God who took on human flesh in order to be one of us, teach us and, ultimately, die for us. This doctrine safeguards us from a Manichean dualism of matter = bad, spirit or mind = good. AI presents an enticing vision of

¹⁸ Jordan Joseph Wales, “Empathy and Instrumentalization: Late Ancient Cultural Critique and the Challenge of Apparently Personal Robots,” in *Culturally Sustainable Social Robotics: Proceedings of Robophilosophy 2020*, ed. Johanna Seibt and Marco Nørskov, Frontiers in Artificial Intelligence and Applications 335 (Amsterdam: IOS Press, 2020), 114–24, <http://doi.org/10.3233/FAIA200906>; David J. Gunkel and Jordan Joseph Wales, “Debate: What Is Personhood in the Age of AI?,” *AI & Society* 36, no. 2 (January 3, 2021): 473–86.

escaping the vicissitudes of the physical, but it is a false vision. The matter that AI is attached to is always there, just hidden. When transhumanists, such as Ray Kurzweil, suggest that we will soon be able to effect our own immortality by uploading our minds to computers, they seem to forget this. A mind in a computer is still operating on a material platform, one that will ultimately fail.

We need to do a better job of teaching the sanctity of the physical world and the importance of our embodiment to our children, who spend so much time in cyberspace, playing video games or on social media, rather than playing in or getting to know the natural world. AI might separate us further from the natural world in which we are embedded and on which we will remain dependent.

Levi Checketts: A very promising result of the rise of AI and its dominance in our culture is the vocal resistance to it as *the* hegemonic concept of intelligence and cognition. Many have raised their voices about the failure of AI to properly account for our embodied nature. Noreen was the first to do this in a theological forum 20 years ago, but we see similar voices in technology studies and philosophy of technology.¹⁹ What these voices remind us is that the idea that humans are primarily *rational* runs the risk of denying that we are also *animal*. This idea finds its logical conclusion in the philosophy of transhumanists like Ray Kurzweil and Martine Rothblatt, who want to totally sever human consciousness from the body through computer uploading.²⁰ Against this, James Keenan notes that Catholic theological anthropology gives priority to the body: we are not merely embodied spirits; we are bodies as much as spirits.²¹ Catholics live an embodied faith: we kneel, embrace, cross, consume, smell, and gaze during Mass. We believe in sacraments—physical manifestations of God’s grace. We revere relics, physical remains of the saints. Above all, we believe that the corpus of the faithful is the mystical body of Christ. The dismissal of the body by AI researchers is a threat to all of this—including the recognition that I am connected, corporeally, to all whom I encounter in partaking in the Eucharist.

¹⁹ See Dreyfus, *What Computers Still Can’t Do*; and Donna Haraway, “A Cyborg Manifesto,” in *Simians, Cyborgs, and Women: The Reinvention of Nature* (New York: Routledge, 1991), 149–82.

²⁰ See Ray Kurzweil, *The Singularity is Near: When Humans Transcend Biology*, PDF e-book ed. (New York: Viking, 2005), 209–20; and Martine Rothblatt, “Mind is Deeper than Matter: Transgenderism, Transhumanism, and the Freedom of Form,” in *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology and Philosophy of the Human Future*, ed. M. More and N. Vita-More (Malden, MA: Wiley-Blackwell, 2013), 317–26.

²¹ James Keenan, SJ, “Roman Catholic Christianity—Embodiment and Relationality: Roman Catholic Concerns about Transhumanist Proposals,” in *Transhumanism and the Body: The World Religions Speak*, ed. C. Mercer and D. F. Maher (New York: Palgrave Macmillan, 2014), 160.

Noreen Herzfeld: Levi, those are good points. Furthermore, we are trying, with AI, to create something in our own image. But that image is partial and distorted. We identify with our minds, which we then consider to be coterminous with our brains. We now know that what we consider to be our “mind” extends into the enteric nervous system and is even influenced by our microbiota. An AI truly “in our image” would need to extend far beyond a simulation or replication of the neural structures of the brain. Any AI not in a biological and mortal body will not exhibit the kind of intelligence or emotion we do. Emotion is a four-stage process. We perceive a stimulus, have a bodily reaction (such as a surge of adrenaline, or of neurochemicals such as dopamine), analyze both stimulus and feeling, and then respond. An AI can perceive a stimulus, analyze it, and respond, but it cannot have a bodily reaction. Its emotional response will, thus, always be somewhat superficial.

Anselm Ramelow, OP: Noreen, building on that, AI does not *have* a body; it *is* a body. It does not “em-body” its procedures, because there are no procedures that it follows: to “follow” a procedure is an act of intentionality, and only beings that have intentionality can be said to “have” a body rather than simply to “be” one. Their intentions are embodied in a physical organism, such that the intentionality becomes its very life. Tools are not alive; they are not part of our bodies even if we become cyborgs. We talk about our tools as if they had intentionality (our computer “seeks” a network, “searches” its files, “tries” to connect with a printer), but this is only the extension of the life with which *we* invest the computer as our tool (*we* are searching the files *with* it). Intentionality itself as one of the basic features of consciousness cannot be accounted for physically, because, as Raymond Tallis notes, it is “causally upstream.”²² Even a basic act of awareness is *directed at* an object from which auditory or visually perceptible waves are emitted. Light waves go one way, our awareness goes the other. Ontologically, this is connected to what Aristotle called “final causality” (in contrast to the “efficient causality” of light waves). Insofar as our nature has a telos, a “final cause,” it intends something, is about something, has a meaning; it unifies that very body as its animating soul. There is no reason to assume any of this for our tools, including AI.

Marga Vega: As Fr. Anselm notes, we find ontological differences and commonalities in the world around us. Exploring that ontological diversity is fruitful. In this regard, AI offers an “ontological” opportunity: the chance to rediscover *who we are as persons* and *what we are* as individuals of the human species. As *persons*, and concerning the AI project, we are more than merely intelligent creatures; we

²² Raymond Tallis, *Aping Mankind: Neuromania, Darwinitis, and the Misrepresentation of Humanity* (Durham, UK: Acumen, 2011), 104–10.

have a relational existence. As *humans*, we are living, physical organisms, so our intelligence is not only naturally sourced; it is embodied. Therefore, one of the *ontological self-discoveries* that AI brings is examining whether intelligence is a sufficient requirement for personhood in the first place.

Cory Labrecque: Building upon what everyone has said, the impact of technology, writ large, on embodiment is of particular interest to me as well and is a subject on which I have written before, especially in the context of religion and transhumanism.²³ The merging of biology and technology (or the technologization/mechanization) of the human body is no longer science fiction: the implantation of microchips in the body, the development of exoskeletons and bionic limbs, designer babies, smart contact lens technology, brain-computer interfaces and neuroprosthetics are just a few examples of the “blurring [of the] perimeter of the body” as TED Fellow and “body architect” Lucy McRae describes it.²⁴ This integration of technology and the body, while not new, requires us to revisit the age-old question that stirred the psalmist who gazed up to the heavens: “What are human beings, O Lord, that you are mindful of them?” (Ps 8:4). More broadly, what are the characteristics of humanhood we must preserve (if any)? Can the human body be modified *ad infinitum* without risking what it means to be human? John Paul II made plain that the human person, who exists as a unity of body and soul (*corpore et anima unus*), is nonetheless a body—that is, “a body among bodies”—rather than merely *having* a body.²⁵ We are, as Kathleen Kalb describes, *body-*

²³ See Cory Andrew Labrecque, “Morphological Freedom and the Rebellion against Human Bodiliness: Notes from the Roman Catholic Tradition,” in *Religion and Transhumanism: The Unknown Future of Human Enhancement*, ed. Calvin Mercer and Tracy J. Trothen (Santa Barbara, CA: Praeger, 2015), 303–13; Cory Andrew Labrecque, “Transhumanism, (Secular) Religion, and the Biotech Age: Liberation from the Lamentable,” in *Everyday Sacred: Religion in Contemporary Quebec*, ed. Hillary Kaell (Montreal and Kingston: McGill-Queen’s University Press, 2017), 234–53; Cory Andrew Labrecque, “Creationism of Another Kind: Integral Corporeality, the Body, and Place in the Catholic Tradition,” *Practical Matters Journal* 9 (2016), wp.me/p6QAmj-FS; Cory Andrew Labrecque, “The Glorified Body: Corporealities in the Catholic Tradition,” *Religions* 8, no. 166 (2017): 1–9; and Cory Andrew Labrecque, “Personhood, Embodiment, and Disability Bioethics in the Healing Narratives of Jesus,” *Journal of Humanities in Rehabilitation* (2017), scholarblogs.emory.edu/journalofhumanitiesinrehabilitation/2017/10/17/personhood-embodiment-and-disability-bioethics-in-thehealing-narratives-of-jesus/.

²⁴ See Lucy McRae, “Compression Cradle,” 2020, www.lucymcrae.net/compression-cradle. See also, for example, Charles E. Binkley, Michael S. Politz, and Brian P. Green, “Who, If Not the FDA, Should Regulate Implantable Brain-Computer Interface Devices?,” *American Medical Association Journal of Ethics* 23, no. 9 (September 2021): 745–49, journalofethics.ama-assn.org/article/who-if-not-fda-should-regulate-implantable-brain-computer-interface-devices/2021-09.

²⁵ John Paul II, *Man and Woman He Created Them: A Theology of the Body*, trans. by M. Waldstein (Boston: Pauline, 2006), 152.

persons who become sacrament in and through the body.²⁶ Although the Church does not outright forbid modification of the body (especially in a healthcare context that strives to preserve and heal), it cautions against a certain sense of “morphological freedom” (to use a transhumanist term) that can threaten corporeal integrity, lead to the absolutization of the body, and promote a cult of the body, as it were.²⁷ In the end, it will be important for us to reflect on the role of technology in replacing bodies or assisting bodies when larger society has chosen to ignore bodies at times.

Anselm Ramelow, OP: Yes, Cory, what Pope John Paul II argues against is a kind of Cartesian dualism. Indeed, our having a body is not like having a car... or a computer, for that matter. But unlike a merely corporeal object, we relate to and have our body. As a consequence, we are not just moved by other objects, but we lead our lives.

Also, I wonder if in emphasizing the embodied aspect of our nature, we are underrating our human distinctiveness from animals. Should we not also defend humans against AI on the basis of human spirituality? Bodies are material, and if anything, computers are material entities—and only that. Just focusing on embodiedness will not make that distinction. It may even reinforce the contemporary “cult of the body” that you mention; and it also leaves angels without their proper status! The importance of the human body (reinforced in the incarnation and the sacraments) has to do specifically with a body that is spiritually animated. What can be done to better spell this out?

Noreen Herzfeld: Jeffrey Pugh has suggested that our fascination with AI and transhumanist goals that consider either uploading our minds to computers or making intelligent computers our progeny represents a return to a Manichaean form of Gnosticism that views the material world as evil and the spiritual/intellectual world as good.²⁸ I certainly do get the sense in reading works by folks like Kurzweil that they think the body is something to be gotten rid of and our identity is coterminous with our brain. This is contradicted by recent work by neuroscientists such as Antonio Damasio who writes that while “any theory that bypasses the nervous system in order to account for the existence of minds and consciousness is destined to failure ... any theory that relies exclusively on the nervous system to account for minds and consciousness is also bound to fail.”²⁹

²⁶ Kathleen A. Kalb, “‘Theology of the Body’ Underpins Health Care,” *Health Progress* 93, no. 2 (March–April 2012): 43.

²⁷ *Catechism of the Catholic Church*, no. 2289.

²⁸ Jeffrey Pugh, “The Disappearing Human: Gnostic Dreams in a Transhumanist World,” in *Religion and the New Technologies*, ed. Noreen Herzfeld (Basel: MDPI, 2017), 51–60.

²⁹ Antonio Damasio, *Feeling and Knowing* (New York: Penguin Random House, 2021), 21.

Cory Labrecque: Some Christian ecotheologians, like Sallie McFague, drawing upon the incarnation and the sacraments which give certain value to the physical world, will say that the resurrected Christ “is present in and to *all* bodies” and that, ultimately, “*all* bodies can serve as ways to God.”³⁰ Being a “body among bodies” emphasizes, at least to some degree, an important commonality and solidarity in our creatureliness (after all, humans and animals alike were made from the dust of the ground, Gen 2:7, 19). There is a deep sense of interrelatedness and interdependence among bodies that cannot, and should not, be cast aside here.

All of this said, it is the human person—a body-soul composite, whose spiritual dimension ought to be understood together with the physical, social, and historical—who alone is created in/as the *imago Dei* for relationship.³¹ Here the distinction between the human-as-body and the non-human-animal-as body (or other bodies for that matter) is made plain, I think.

Moderators: Moving from body to mind, how might AI technology enlighten (or complicate) theological and philosophical perspectives on the meaning of intelligence and consciousness?

Andrea Vicini, SJ: One wonders whether “intelligence” is the most appropriate term to describe algorithmic computation and analysis. The term “artificial intelligence” is so commonly and widely used that it is pointless to even consider proposing to replace it. Still, I would prefer to reserve “intelligence” for the unique and, until now, unmatched abilities of human intelligence, with all its strengths and limitations.

Jordan Joseph Wales: This is an important point to explore. “Artificial intelligence” began as a reflection of mid-century self-understandings; now—for ill—it sometimes is taken more as a defining point of reference than as a reflection.

A thousand years ago, *intellectus* meant the intuitive grasp of something as it is in itself; *intellectus* was the clear vision underlying all discursive reasoning.³² In the 1950s, AI meant the computational accomplishment of feats that would ordinarily require human thinking and insight: planning, chess-playing, etc. In the 1980s, as robotics became more popular, the logicist reduction of *intellectus* to

³⁰ Sallie McFague, “The Scope of the Body: The Cosmic Christ,” in *This Sacred Earth: Religion, Nature, Environment*, 2nd ed., ed. Roger S. Gottlieb (New York: Routledge, 2004), at 262 and 266.

³¹ International Theological Commission, “Communion and Stewardship: Human Persons Created in the Image of God,” 2004, 1.9–10, www.vatican.va/roman_curia/congregations/cfaith/cti_documents/rc_con_cfaith_doc_20040723_communion-stewardship_en.html.

³² Josef Pieper, *Leisure: The Basis of Culture* (San Francisco: Ignatius, 2009).

computation was followed by a further reduction of the AI to a “rational agent” that “acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome.”³³ Here, then, with historian Yuval Noah Harari, we may re-describe “intelligence” as “the ability to solve problems.”³⁴ Projected onto humans, this approach reduces us to the “instrumentalized reasoning” that Charles Taylor and Alasdair MacIntyre identify as characteristic of our age. If our intelligent machines are intelligent in behaving so as to fulfill our purposes, then are our neighbors also intelligent insofar as they conform to our purposes? Under such a view, Taylor writes, all things are “open to being treated as raw materials or instruments for our projects.”³⁵

This, of course, is the pride that Augustine considers to be the root and deepest outcome of the fall. The reduction of intelligence to logic, and then to behavior—without reference to an interior life—risks shifting our cultural language so as to depict human life as a task of optimizing (my) benefit, to the exclusion of mutual self-gift. At the limit, we may come to see one another (and even ourselves) simply as behavior-producers, whose value will be quantifiable in terms of the production of desired actions. With the recent rise in the tracking of personal activities, habits, fitness, and performance—despite the obvious benefits of these technologies—we may see this shift already in progress.

Paul Scherz: I really like how Jordan provides a historical outline of the understanding of intelligence in AI. However, I wonder if we have not moved on to a fourth stage beyond the movement from classical *intellectus*, to computation, to instrumental reason. With contemporary forms of machine learning, as they are being deployed across the economy and government, the goal seems merely to make predictions, things like the behavioral futures that Shoshanna Zuboff discusses.³⁶ Intelligence becomes something akin to gambling skill.

What I find interesting is how models of intelligence used in the programming realm feed back into areas of human activity. As I and others have noted, the widespread use of machine learning and Big Data is transforming many fields of science such as genetics, which is

³³ Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. (Upper Saddle River, NJ: Pearson, 2009), 2.

³⁴ David Kaufman and Yuval Noah Harari, “Watch Out Workers, Algorithms Are Coming to Replace You—Maybe,” *New York Times*, October 18, 2018, www.nytimes.com/2018/10/18/business/q-and-a-yuval-harari.html.

³⁵ Charles Taylor, *The Ethics of Authenticity* (Cambridge, MA: Harvard University Press, 1991), 5.

³⁶ Shoshanna Zuboff, *The Age of Surveillance Capitalism* (New York: PublicAffairs, 2019).

flooded with genomic data.³⁷ There comes to be an expectation that scientific knowledge and discovery will come merely from having machines churn through ever larger piles of data, with some even suggesting that AI systems will be able to perform their own research in an era of “hypothesis-free” research. The ways these changed concepts of intelligence and understanding filter back into scientific practice are causing significant distortions in contemporary research. I would imagine that these kinds of effects are being seen in a number of fields.

Brian Cutter: Very interesting thoughts, Jordan. And, like Paul earlier quoted, I am reminded of one of my favorite passages from C. S. Lewis’s *The Abolition of Man*:

If man chooses to treat himself as raw material, raw material he will be: not raw material to be manipulated, as he fondly imagined, by himself, but by mere appetite, that is, mere Nature, in the person of his dehumanized Conditioners. ... Either we are rational spirit obliged for ever to obey the absolute values of the *Tao* [natural law], or else we are mere nature to be kneaded and cut into new shapes for the pleasures of masters who must, by hypothesis, have no motive but their own “natural” impulses.³⁸

Jordan Joseph Wales: Thank you, Brian. From a similar time period as Lewis, we might also cite Winston Churchill, who believed that not the British but the Soviet society would be best suited for robotic slaves because they would be the final fulfillment of what Churchill saw as the Soviet view of the person as a cog in the machinery of state. But now we find the same view attributable to tendencies in our own society, as Lewis foresaw.³⁹

Marga Vega: Jordan, you make an important point about the history of AI research. Under the computational theory of the mind and cognitivism, the first years of artificial intelligence encouraged computer scientists’ hope to achieve machines that could think not just *like* humans but also *better than* humans, possibly even showing consciousness. Conversely, it also opened the prospect of mapping the human mind in computational terms, dismissing the importance of consciousness and awareness in cognition, and leveling any assumed

³⁷ Paul Scherz, “The Displacement of Human Judgment in Science: The Problems of Biomedical Research in an Age of Big Data,” *Social Research* 86, no. 4 (2019): 957–76; Erik Larsen, *The Myth of Artificial Intelligence: Why Computers Can’t Think the Way We Do* (Cambridge, MA: Belknap, 2021); Jenny Reardon, *The Postgenomic Condition* (Chicago: University of Chicago Press, 2017); and Hallam Stevens, *Life Out of Sequence* (Chicago: University of Chicago Press, 2013).

³⁸ Lewis, *The Abolition of Man*, 73.

³⁹ See Churchill’s essay “Fifty Years Hence,” in *Thoughts and Adventures: Churchill Reflects on Spies, Cartoons, Flying, and the Future*, ed. James W. Muller, Paul H. Courtenay, and Alana L. Barton (Wilmington, DE: ISI, 2009).

ontological differences between the human mind and machine intelligence.

Underlying both projects is the analogy that the mind is to software as the brain is to hardware. How far we take this analogical seesaw by conceding more weight to the idea that machines have minds or to the idea that human minds are machines may not matter much if the result in both cases is to minimize the ontological differences between minds and machines under a paradigm that equates computation with intelligence.

The problem with equating computation and intelligence is that computation is only possible if there are minds relative to which we can assign computational interpretations. In other words, computation cannot ground intelligence because intelligence is an *a priori* condition for computation. John Searle's Chinese Room Argument (CRA), which initially pointed at the lack of semantics in computers, later addressed this difficulty with the thesis "syntax is not physics."⁴⁰ In comparing minds and computers, the CRA noted that it is not only that computers have a syntax and not semantics; they do not even have a syntax since any syntactical structure is observer relative. Syntax exists only relative to minds capable of mental content, and that is precisely what is at stake in the case of computers: the capacity to have something other than purely physical causal processes devoid of mental content.

Jordan Joseph Wales: Marga, I like the way you are going here. Even before we speak of a soul, we must ask whether the chemical reactions in the nervous system have some causality beyond that which is describable in physics and chemistry. If physics and chemistry as presently understood exhaustively describe our bodily processes, then there is neither consciousness nor meaning—a claim that seems manifestly false by our very experience.

Anselm Ramelow, OP: Indeed. When we talk about ourselves, what we mean by "consciousness" has features we do not expect machines to have, among them a subjectively experienced point of view, intentionality and, for rational minds, a kind of reflexivity that cannot be instantiated in material objects.⁴¹ Another feature is a certain

⁴⁰ Searle introduced the idea of syntax being observer-relative in his Presidential Address to the American Philosophical Association, and it has appeared since in subsequent formulations of the CRA such as "Who is Computing with the Brain," *Behavioral and Brain Sciences* 13, no. 4 (December 1990): 632–42, and *The Mystery of Consciousness* (London: Granta, 1997). Sometime after 2003, the argument appears as the "syntax is not physics" thesis in Searle's lecturing and writings.

⁴¹ A point also made by Karl Rahner, SJ, "Person. II. Man. C: Theological," in *Sacramentum Mundi: An Encyclopedia of Theology*, vol. 4: Matter to Phenomenology, ed. K. Rahner (Montreal: Palm, 1969), 417. Reflexivity is also the root of creativity; see Anselm Ramelow, OP, "Can Computers Create?," *Evangelization and Culture* 1 (2019): 39–46.

“unified” character: consciousness is a unifier of all its contents. Ontologically, this feature corresponds to the unified life of organisms, which differentiate themselves from their environment both in their actions and in their very being. Characteristically, these are kinds of unities we cannot make ourselves. Living beings originate by procreation, not artificially (*omne vivum ex vivo*). Why would we expect this to be different in the case of consciousness, which is a life that has the additional unifier of awareness? The making of conscious entities may require, therefore, the causality of someone who gives things both their nature and existence, the most fundamental unifying properties. Such a maker would therefore need to be a creator (God). We, on the other hand, *presuppose* the existence of things and *rely on* their natures in order to build artifacts with them. These artifacts do not have any other unity than the purpose we have for them. The unity is not ontological or intrinsic to them, but only in our minds. This is true for AI as well: neither in its being nor in its operations does AI have the requisite unity to be conscious. Metaphysically, the parts are in potency with regard to the whole; hence the actualization of this unified whole requires a proportionate cause. If the unity in question concerns the very nature of the thing, this cause may need to be a creator.

Marga Vega: That is a relevant question, Fr. Anselm, whether consciousness can pertain to entities that are not alive, and whether consciousness is itself a type of life. If the latter happens to be the case, it seems that a conscious artifact is not possible. However, some would defend the proposition that perhaps intelligent computers do not need consciousness—all that is required is intelligent behavior. It is questionable that what is meant by “intelligence” in the case of humans and in that of computers can be taken univocally.

But even if we generously granted “intelligence” to computers, their status as artifacts and non-persons would remain. Even for those unfamiliar with Boethius’s definition of the person as an individual substance of rational nature, the idea that rationality grounds our personhood takes hold of our minds both through our civilization’s history and our personal and societal values. Based on this intuition, some have questioned, with perplexity, the personhood of human beings whose rationality is impaired. Embryos, neonates, people in vegetative state, or those with disabilities may lack the exercise of intellectual capacities that some would consider essential for personhood.

Likewise, based on intelligence, a debate emerges on whether machines could have, if not metaphysically then at least legally, the status of persons. If we have machines that compete with us in terms of intelligence, should they also qualify as persons if intelligence characterizes personhood? The paradox is that placing intelligence as the paragon for personhood may strip the title of “person” from humans with dormant rational capacities while entertaining whether machines could be eligible candidates for this status. The challenge of AI offers

us an ontological opportunity: perhaps intelligence is not a definitive measure for personhood, if personhood (or even humanity) does not ensue from the possession of an ability. On the contrary, personhood precedes any capacity.

Jordan Joseph Wales: Marga, I understand wanting to uphold the personhood even of those who have dormant faculties, but you make it seem as if denying personhood to machines is a foregone conclusion.

Marga Vega: We tend to infer what something is from the way it acts. At first sight, it would seem that: (1) if a computer acts intelligently, then it is intelligent; and therefore (2) it can be counted as a person. From the point of view of how we get to know things, this would seem like a valid inference. But we must not confuse epistemology with ontology, how things are.

First, behavior alone does not guarantee that what causes the behavior is the same in both cases. A sore throat may be a sign of the flu but also of COVID-19. Performing specific intelligent tasks may have a comparable output by a computer and a human, but the causal elements could be very different. Therefore, we cannot conclude intelligence from the appearance of intelligent behavior: we need independent definitions of what counts as intelligence and what kind of causality it requires.

Second, it is questionable whether intelligence or rationality constitutes persons (granted that rationality accompanies personhood). It could well be that rationality does not ensure personhood and that personhood causes rationality. In this case, we would have things backward in assigning personhood to computers based on their intelligence. Therefore, we would need to inquire into what is the root of personhood in the first place.

Cory Labrecque: I think Marga raises an important point here that brings to the fore contemporary wrestling with the definition of personhood. The concept is at the center of bioethical discourse, but so few agree on how it should be understood.

In a short piece entitled “Is Koko a Person?” James W. Walters—Professor of Ethics at Loma Linda University—makes a distinction (well known by theorists who study moral status) between what he calls physicalism and personalism (not to be confused with other uses of this term in philosophy and theology). The former argues that “the essence of a person is found in his or her biological make-up. All humans are persons, ipso facto.” The latter, which is telling here and links to Marga’s critiques, locates the essence of a person “in one’s mental capacities and ability to use these in satisfying ways. Whether one is a human is not important.” In this way, robots and computers could fit the bill, while certain human beings—bereft of certain functions—fall short. This linking of capacity, function, and performance

to dignity and value further deprecates other shared dimensions of humankind: fragility and vulnerability.⁴²

Noreen Herzfeld: I think it is worth noting that “person” has become a legal category here in the US. We allow corporations to be classed as persons, in this legal sense. One problem with “personhood” is that it is a binary—one either is or is not a person. In legal terms, it must be binary, however this makes it less than useful as a philosophical designation. With respect to AI, the fetus, or the severely disabled, I think we would do better to speak in shades of gray, rather than black or white.

Brian Cutter: While I suspect that AI technology will not teach us much about the nature of consciousness, I do want to say there is a lot here we probably cannot really know. If we eventually create an AI that passes behavioral tests for general intelligence (e.g., a Turing test), we probably will not know whether it is conscious, even if it says it is.

In my view, consciousness (i.e., subjective experience) is ontologically distinct from any set of physical or computational processes, so even if we had complete knowledge of the machine’s physical operations, this would not conclusively settle whether it was conscious.⁴³ While consciousness is distinct from any purely physical process, conscious states are obviously *correlated* with certain physical processes (e.g., processes in human brains) in regular, lawful ways. To figure out whether an advanced AI would be conscious, a key question is whether the “psychophysical laws” (the laws of nature by which physical states are linked to states of consciousness) are *substrate-independent*—that is, whether they are sensitive to the material composition of a physical system, or whether they are only sensitive to the higher-level causal organization of the system, abstracting away from its material substrate.

In principle, the high-level causal organization of a human brain could be implemented in a computer. For example, a detailed computer simulation of a human brain would exhibit the same causal organization as a human brain, but it would be realized in a silicon-based material substrate rather than a carbon-based substrate. If the psychophysical laws are substrate-independent, as some philosophers have argued, then a detailed computer simulation of a human brain would

⁴² See James W. Walters, “Is Koko a Person?,” *Dialogue* 9, no. 2 (1997), circle.adventist.org/files/CD2008/CD2/dialogue/articles/09_2_walters_e.htm.

⁴³ I will not defend the ontological distinctness claim here; I accept it on the basis of arguments like those given in David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press, 1996); Brian Cutter, “The Modal Argument Improved,” *Analysis* 80, no. 4 (2021): 629–39; Saul A. Kripke, *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980); and Adam Pautz, “Do Theories of Consciousness Rest on a Mistake?,” *Philosophical Issues* 20, no. 1 (2010): 333–67.

be conscious.⁴⁴ My own view is that we currently have no idea whether the psychophysical laws are substrate-independent, and I doubt this question will be settled any time soon.

Andrea Vicini, SJ: As a final comment on this question, I would just like to point out that concepts such as intelligence, embodiment, natural/artificial, and consciousness do not exist in a vacuum. Each of these concepts tries to articulate a particular dimension of the complex human and social reality with its plurality.⁴⁵ Moreover, how one understands these concepts depends on the historical and cultural contexts in which they are articulated. Critical reasoning should help us examine the elements that characterize our context and how this context influences our understanding of each one of these concepts.⁴⁶ Asking whether or not they promote the common good could help to discern between the various interpretations of these multiple concepts and the concrete implementations they make possible. At the moment, sadly, AI seems to reinforce the social inequities, discriminations, and biases present in our society.

Moderators: Fr. Andrea, this is a good segue to our next point. For each of you, what are some key ethical issues to focus upon with respect to AI? How might Christians and the Catholic Church in particular helpfully respond to these issues?

Noreen Herzfeld: Well, as a first point, I would refer to my article in this issue on lethal autonomous weapons systems (LAWS) and just war theory.⁴⁷ We need to push for bans on LAWS before they become widespread. Should these weapons ever be designed to act with complete autonomy (no human in the loop) the likelihood of unforeseen consequences would be staggeringly high, as would the likelihood of these weapons being deployed by rogue actors. While no ban is totally enforceable, international condemnation does have an effect, as we have seen with chemical weapons. I find it both interesting and heartening that most military and former military generals I have spoken to

⁴⁴ See, e.g., Chalmers, *The Conscious Mind*.

⁴⁵ See Peter G. Kirchschlaeger, "Artificial Intelligence and the Complexity of Ethics," *Asian Horizons* 14, no. 3 (2020): 587–600.

⁴⁶ See Paolo Benanti, "Algor-éthique: intelligence artificielle et réflexion éthique," *Revue d'éthique et de théologie morale* 3, no. 307 (2020): 93–110. See also Paolo Benanti, *Digital Age: Teoria del Cambio d'Epoca: Persona, Famiglia e Società* (Cinisello Balsamo: San Paolo, 2020); Paolo Benanti, *Realtà Sintetica: Dall'Aspirina alla Vita: Come Ricreare il Mondo?* (Roma: Castelvecchi, 2018); Paolo Benanti, *Le Macchine Sapienti* (Bologna: Marietti, 2018); Paolo Benanti, *Oracoli: Tra Algorética e Algorazia* (Roma: Luca Sossella, 2018).

⁴⁷ Noreen Herzfeld, "Can Lethal Autonomous Weapons Be Just?," *Journal of Moral Theology* 11, Special Issue 1 (2022): 70–86.

are adamantly against the design or deployment of fully autonomous weapons.⁴⁸

Levi Checketts: Another pressing problem is the displacement of laborers and the increase of wealth disparity across the globe. One of the most immediate practical uses of AI involves cost-saving and labor-saving procedures. The ultimate result of this will be that those who own or control AI will become fabulously wealthy while the vast majority of others will find themselves competing against a labor system that does not need housing, time off, or provision for biological necessities. Magisterial Catholic social teaching, such as *Laborem Exercens* or *Pacem in Terris*, reminds us that work is a human good and governments have an obligation to ensure the common good above the amassing of wealth (*Laborem Exercens*, no. 17; *Pacem in Terris*, nos. 56, 121). Here, the Church has a vast treasure of resources to turn to, prophetic voices condemning the unthinking use of power to the benefit of few and detriment of many. We might think of the witnesses of liberation theologians, St. Ambrose of Milan or John Chrysostom, and speak out against the unjust use of power and wealth against the poor.⁴⁹ AI should be used for *all* humanity, not only the rich.

Paul Scherz: Levi, I agree that the way AI applications support concentrations of economic and political power is a huge problem. These applications require immense datasets and computing power, so they can be deployed only by large corporations, governments, or other entities with the requisite funding and access to those resources. In workplace settings, AI can be implemented in ways that support deskilling and the centralization of knowledge in management, thus continuing the trend of worker disempowerment seen in Taylorism.⁵⁰ Concentration of power has long been considered a problem in Catholic social thought, insofar as it increases social struggles and decreases the possibility for free action.⁵¹ We see this anew in the way these forms of concentrated power can disempower workers, undermine privacy, and expand bias.

⁴⁸ See Noreen Herzfeld and Robert H. Latiff, "Can Lethal Autonomous Weapons be Just?," *Peace Review* 33, no. 2 (2021): 213–19.

⁴⁹ See, for example, John Chrysostom, *Homily 7 on Colossians*; Thomas Aquinas, *Summa theologiae*, IIa IIae, q. 66, a. 1; and Gustavo Gutiérrez, *A Theology of Liberation: History, Politics and Salvation*, trans. Caridad Inda and John Eagleson (Maryknoll, NY: Orbis, 1973), 163–64.

⁵⁰ For Taylorism and a more general account of the degrading effects of routinization on workers, see Harry Braverman, *Labor and Monopoly Capitalism: The Degradation of Work in the Twentieth Century* (New York: Monthly Review, 1998). For broad accounts of deskilling in the wake of automation, see Nicholas Carr, *The Glass Cage* (New York: Norton, 2014); and Shannon Vallor, "Moral Deskilling and Upskilling in a New Machine Age," *Philosophy and Technology* 28, no. 1 (2015): 107–24.

⁵¹ E.g., Pius XI, *Quadragesimo anno: On the Reconstruction of the Social Order* (1931), nos. 105–109, www.vatican.va/content/pius-xi/en/encyclicals/documents/hf_p-xi_enc_19310515_quadragesimo-anno.html.

Andrea Vicini, SJ: Agreeing with Levi and Paul, threats to social justice, including racial discrimination and increasing inequities, are relevant ethical challenges.⁵² Technological progress fostered by AI ought to promote greater equality by addressing the increasing gap between those who have and those at the margins of the social fabric. AI should not further heighten social inequities. Christians and Catholics, both as individuals and ecclesial institutions, share social responsibility towards fostering awareness on the part of citizens and believers regarding uses of AI technology that disempower and marginalize, and to join multiple social actors (e.g., citizens, groups, and organizations—nationally and internationally) in addressing these diverse ethical challenges in collaborative ways with scientists, politicians, activists, communities, and multinational companies. Finally, with a variety of its agencies and the leadership of Pope Francis, the Vatican appears to be at the forefront of dialogue, reflection, and critical engagement regarding AI involving scientists, scholars in the humanities, universities, and biotech companies.⁵³ Such an engagement is praiseworthy and shows how it is possible, even necessary, to be participants in the social arena by joining multiple social forces while aiming at promoting a broad social agenda, open toward progress and the future, and animated by a realistic hope.

Cory Labrecque: The issues are myriad. One topic that does not often come to the fore is the importance of touch for healing, and how the transfer of care (or even parts of the care process) to AI software or machines may very well suppress crucial bodily elements of the patient-healthcare provider relationship (that is, of human bodies in relationship) conducive to well-being. In her *Broken Nature* exhibit, sci-fi artist Lucy McRae introduces a new work that she calls a “compression cradle” as a response to the “touch crisis” in which we find ourselves due to the lack of physical contact that has come about from our excessive connection with technology. Yet McRae responds to this mark of rampant technologization by introducing yet another technology: a machine that “affectionately” squeezes the body through a series of aerated membranes that “hold you tight in an attempt to prepare the self for a future that assumes a lack of human touch.”⁵⁴

As another point, for the Church, technology must have the good of human beings and the whole human family at its heart. It must be an expression of stewardship and service, contribute to genuine progress (that is, a progress that will lead human beings to exercise a

⁵² As an example, see Isabel Wilkerson, *Caste: The Origins of Our Discontent* (New York: Random House, 2020).

⁵³ See Vincenzo Paglia and Renzo Pegoraro, *The “Good” Algorithm? Artificial Intelligence: Ethics, Law, Health*, Proceedings of the XXVI General Assembly of the Pontifical Academy for Life (Rome: Pliniana, 2021).

⁵⁴ See Lucy McRae, *Compression Cradle* (2020), www.lucymcrae.net/compression-cradle.

wider solidarity and opening themselves more freely to others and to God) (*Octogesima Adveniens*, no. 41), respect the inherent dignity of human beings and all natural environments, and recognize the delicate complexity of ecosystems and the interdependencies extant within them. Although technology may very well extend our dominion over the material world (*Caritas in Veritate*, no. 69), the Church reminds us that the Biblical mandate to subdue the earth and have dominion over it is an *entrusted* one that should never become despotism or absolute mastery/lordship over the body (one's own or others' bodies). The mandate is very much a collective responsibility to make manifest God's love for the whole of Creation. The Church—for whom the corporal works of mercy (i.e., feeding the hungry, tending to the ill, clothing the naked, sheltering the homeless, etc.) shape the Christian moral life as an extension of God's compassion—ought to be on the front line countering these trends.

Marga Vega: Going in a different direction, I think there is a real risk in the delegation of moral decision making to artifacts. Rationality is a capacity that only makes sense as an ability *from* and *for* the ability's owner. Rationality is not an absolute and cannot be uprooted from a teleology nor cut off from someone who holds that rationality. The unique challenge AI poses is what to do when the tool becomes the wielder. Might the wielder then become a tool in the service of the new wielder: a machine? Our created AI could become so self-sustaining and independent that it could hand out decisions leaving us in the dark as to the *criteria* guiding the reasoning process and powerless to resolve what is best. We do not need to jump into a self-driving car to envision scenarios where the bliss of ignorance and referred decision-making can become a liability. It is clear then that we cannot assemble an artificial intelligence without built-in values that guide and preserve personal rational criteria. More than ever, ethics is necessary for technology, not just as *how to decide the proper use* of the technological invention, but *how to build-in values* in the very fabric of the tool's constitution.

Jordan Joseph Wales: Taking the artificial agent question in a different direction again, I have already described the quandary of ethical formation that will arise from our owning the services of apparent but unreal persons. Several of us have also attended to how our beliefs concerning AI may reshape or distort our understanding of the human person. Therefore, Christians and the Church must above all bear witness to the importance of the human interior life and to the self-gift that flows from it. Even before and beyond Christians' and the Church's declarations concerning AI, it is this relational witness that will preserve in our culture the means by which to live humanly alongside and with the fantastic technological developments of today and tomorrow.

Noreen Herzfeld: Building on that, one place where self-gift is most evident is in our sexual lives. In the act of intercourse, we give our bodies to another in an openness and vulnerability that ideally flows from the interior life Jordan speaks of. Increasingly lifelike robots, or sexbots, are already being developed to function as our electronic lovers. While these might make interesting or even desirable sexual partners, they represent another form of idolatry, substituting a relationship with the living with something made, and thus controlled, by our own hands. In this way we risk reducing sex to a one-way street, in which the robot is there to meet our needs and proclivities, emptying the act of its wildness and mystery and making few demands upon us. It becomes a form of whoredom.⁵⁵

Brian Cutter: For me, the most philosophically interesting ethical question about AI is whether an advanced AI would itself have moral status—whether it would have morally significant interests we ought to respect (my concern here is with hypothetical future AI, not current AI). This would partly depend on whether it is conscious (i.e., capable of subjective experience). A capacity for subjective experiences like pleasure and pain is, I think, a sufficient condition for having *some* moral status, though not a sufficient condition for the *full* moral status associated with persons. Thus, even if an advanced AI with the right cognitive architecture would be conscious, and therefore have some moral status, it might not have full moral status. We should also think about how to navigate the issue of *moral risk* in this domain. How should we treat an advanced AI if we are unsure whether it has any moral status (say, because we are unsure whether it is conscious?). It would be interesting to explore the analogies and disanalogies to the issue of moral risk in debates about the ethics of abortion. A common pro-life argument is that abortion is gravely wrong even if we are unsure whether a fetus is a person, since in general, it is wrong to do things that carry a significant risk of killing an innocent person.

Anselm Ramelow, OP: Brian, the “moral risk” question is interesting. Still, there are two disanalogies here: (1) in the one case there is a high-performance entity where we are unsure of consciousness, in the other we do assume neither consciousness nor expect high performance yet think mistreatment risky. Why? (2) The causal history of both entities is different: one we have made, the other begotten. Life comes only from life, should consciousness not come a fortiori from conscious beings? Maybe that answers also point (1)?

Brian Cutter: Fr. Anselm - Interesting points! I agree there is some disanalogy. Most importantly, the “risk” is much greater in the abortion case, since the fetus is a member of the human species, and it

⁵⁵ See Noreen Herzfeld, “Religious Perspectives on Sex with Robots,” in *Robot Sex: Social and Ethical Implications*, ed. John Danaher and Neil McArthur (Cambridge, MA: MIT Press, 2017), 91–102.

is extremely plausible that being a member of the human species is a sufficient condition for full moral status. Pro-choice philosophers typically reject the latter claim on the grounds that the fetus, though a human being, does not have sufficiently developed mental capacities to qualify as a “person.” That objection has implausible and repugnant implications, e.g., infants and the severely mentally disabled lack full moral status. I do not think the principle that conscious beings only come from conscious beings settles the question. AI comes from us, and we are conscious. Here I assume “comes from” covers creating and not just begetting; otherwise, the principle that “life only comes from life” is false, since the first life forms were not begotten by any living thing.

Anselm Ramelow, OP: Yes, indeed. Here, further arguments would need to be made that the first making of life can only be done by a Creator. We at least cannot; and to me that seems to be an important difference.

There is an opportunity for Catholic philosophy and theology to speak on issues related to AI. Contemporary thought is hesitant to reflect on the whole of reality. Metaphysical questions or questions of the meaning of life do not typically receive robust answers. The Catholic Church is, for theological reasons, more confident in the ability of our mind to propose such answers, and she has a long history of formulating such answers. It is time to take such proposals from the shelf and articulate them anew. This need not require an attempt to prove the truth of these proposals. It may rather be proposed as an “inference to the best explanatory hypothesis.” If it gives a richer and more cogent explanation of reality, including of humans and AI, and accounts for more data, including for our moral intuitions, then it has the chance to be helpful as a response.

Moderators: Fr. Anselm, that is a great transition to our last question about the practical relevance of these questions about AI. What can and should Catholic institutions such as universities, hospitals, charities, and the Vatican, and Christian institutions more broadly, do in order to facilitate the better uses and restrict the worse uses of AI?

Jordan Joseph Wales: Catholic institutions must become well educated as to how AI works and reflect deeply both on AI and the human person in order. So doing, both individuals and institutions can, as members of the ecclesial body, advocate appropriately, for instance, for laws that prohibit the mistreatment of apparent persons (on the basis not of personal rights, which they cannot have, but of their signifying of the personal, much like public anthropomorphic works of art). Persons of any faith or none should remind themselves again and again, that the instrumental functioning within which we use these

tools does not exhaust the meaning of the human person. However, that is not all that we must say; we must consider also how the theology of creation and the human person allows us to think more carefully about just *what* this or that AI might be. The theory of biological evolution invites theological reflection on an unfolding divine providence in light of ancient Christian beliefs concerning the manifestation of God's wisdom in the created order. So too, the rise of artificial intelligence—especially deep learning and its capacity to apprehend hidden dynamics within large data sets—allows us to think again about the ways in which machines designed for our purposes can be both attuned to the deep dynamics of contingent events (e.g., markets, societies, and the weather) and can also obscure those dynamics (e.g., in AI bias) depending on how we have carved up the reality we seek to engage. Human engagement with the world is, from a Catholic point of view, a theological and a spiritual phenomenon; the more Catholic institutions and theologians reflect on artificial intelligence, the more—and the more usefully—we shall find we have something to say.

Andrea Vicini, SJ: Institutions of higher education play an important social role with their teaching and research. They contribute to the education, formation, and training of students and citizens in engaging the technological developments and social implementations of AI in critical ways, in light of an articulated ethical approach, with a strong attention given to social dynamics and their historical implementation.⁵⁶ For example, the history of medicine, technology, and science allows us to learn both from virtuous and vicious approaches by considering, respectively, benefits and troubling consequences and addressing any injustice and inequity.

Paul Scherz: I agree with Andrea's point as to the institutional importance of Catholic colleges and universities. They can act as both important research centers for exploring these questions as well as centers for forming the next generation of citizens in using these technologies well. In the latter role, Catholic institutions of higher education can serve as a crucial witness as to how to embody the use of the technologies well, or as a prophetic witness as to what instances and uses of these technologies must be rejected.

Catholic health care is even closer to the front lines on these issues. These health systems have vast troves of data on their patients, and technology companies are eager to gain access through partnerships that will allow them to sift through the data with their machine learning systems. These health systems must be careful about, on the one hand, falling to the hype about the promise of artificial intelligence and thus overpromising what these partnerships might achieve and, on the other hand, using their patients' data in exploitative ways. The

⁵⁶ See Angelo Chakkanattu, CMI, "Artificial Intelligence: Human Natural Machine Intelligence of Evolution," *Asian Horizons* 14, no. 3 (2020): 563–86.

structure and implementation of such partnerships will help determine whether these programs respect important goods such as dignity, privacy, the common good, and service to the poor, or whether they make use of the data solely for instrumental goods of profit or, worse, a biased delivery of healthcare. Even with the best intentions, a poorly structured program could lead to dangerous practical effects. This is a place where moral theologians along with scholars from other fields such as law could, in a very practical manner, help these systems fulfill their role as a ministry of the Church. Scholars can assist in examining how to prevent the dangers and encourage the positive potential of these partnerships. Healthcare is another area in which Catholic institutions could be a witness as to how to use these technologies well.


Levi Checketts: One risk I see is people rushing headlong into the technology, seemingly pursuing what Max More calls the “proactionary principle”—focusing on developing technology first and adjusting it to the good later.⁵⁷ I have been party to many conversations where technologists aptly demonstrate their knowledge of the field of AI and how *they* think it might be good, but they often think their opinion of the moral problems surpasses the understanding of theologians and philosophers. One reason why this problem exists, in my view, is that theologians involved in interdisciplinary discussion cut right to engineers and entrepreneurs as dialogue partners, rather than dialoguing with technology scholars, policy advisors, social theorists, philosophers of technology, and critical theorists. We need to have more events to discuss theology and AI, but we need to open the forum so that scholars, activists, and ministers who are not inherently engaged with AI production are discussing it. In my research on transhumanism, with dozens of responses from theological thinkers, one of the single best responses I read was that of James Keenan, who has no interest whatsoever in transhumanism. As an outside scholar, he was able to articulate problems that many too close to the issue had missed, such as the nature of Catholic collectivism and embodiment. Likewise, if we invite feminist theologians, ecological theologians, critical theorists, black, Latinx, Asian theologians, and others into the dialogue, we may find creative responses and ideas for the problems at hand.⁵⁸

Noreen Herzfeld: Levi makes a good point. We need to remember that most AI research is funded with soft money. This means the researchers must hype the possible good outcomes that will come out of it. As one technologist at MIT put it “We shall overclaim!” Yet we cannot have all the critique coming from the outside. Many of the ethical dilemmas that appear in our technology are baked into the design. Thus, it is imperative that designers themselves start asking, not only

⁵⁷ Max More, “The Proactionary Principle: Optimizing Technological Outcomes,” in *The Transhumanist Reader*, 258–67.

⁵⁸ Keenan, “Roman Catholic Christianity—Embodiment and Relationality, 155–72.

what good their product can do, but also what harm. When harm occurs, the corporations that design, market, and run our computer systems need to take responsibility. Ultimately, a computer, as a non-sentient thing, remains a tool. It cannot be a moral agent. Only humans are that.

Moderators: Thank you, Noreen; on that note we have run out of time. This has been a great conversation! Thank you to all of you for your contributions. 

Brian Patrick Green is Director of Technology Ethics at the Markkula Center for Applied Ethics at Santa Clara University; Matthew J. Gaudet is Lecturer of Engineering Ethics at Santa Clara University and Fellow at the Grefenstette Center for Ethics in Science, Technology, and the Law at Duquesne University; Levi Checketts is Assistant Professor of Religion and Philosophy at Hong Kong Baptist University; Brian Cutter is Associate Professor of Philosophy at the University of Notre Dame; Noreen Herzfeld is the Nicholas and Bernice Reuter Professor of Science and Religion at St. John's University and the College of St. Benedict and senior research associate with ZRS Koper; Cory Andrew Labrecque is Associate Professor of Bioethics and Theological Ethics, and the inaugural Chair of Educational Leadership in the Ethics of Life at the Faculty of Theology and Religious Studies at Laval University; Anselm Ramelow, OP, is Professor of Philosophy at the Dominican School of Philosophy and Theology; Paul Scherz is Associate Professor of Moral Theology and Ethics at the Catholic University of America; Marga Vega is Professor of Philosophy at the Dominican School of Philosophy and Theology; Andrea Vicini, SJ, is Michael P. Walsh Professor of Bioethics and Professor of Theological Ethics in the Boston College Theology Department; and Jordan Joseph Wales is Associate Professor and the John and Helen Kuczmariski Chair in Theology at Hillsdale College.