Artificial Intelligence

Can we ever consider robots, computers, or machines people? What is the current state of the field in artificial intelligence? What are the current challenges in creating artificial intelligence and how can that develop our understanding of intelligence and life? And what sort of criteria might we use to evaluate whether AI is "alive" or has personhood?

ARTIFICIAL INTELLIGENCE - I

Can Robots Be Persons?

— ANNE FOERST —
PROFESSOR OF COMPUTER SCIENCE, ST. BONAVENTURE UNIVERSITY

W HEN WE TALK about something being alive, we are talking about a biological category. This category is value-free in and of itself. We do not automatically deem everything that is alive as intrinsically valuable, as the common use of antibiotics shows.

Similarly, when looking at the current crop of intelligence machines, we do not assume they have intrinsic value; we question whether these robots *can* have intrinsic value. We will first look at some current machines and then address that particular issue.

AI-driven robots and other machines have become better and more tactile and begun replacing human beings to perform many menial tasks. Especially in harvesting, the progress has been rapid; some produce, like grapes, that was always harvested by hand can now be harvested by robots. But machines have been used in factories and agriculture since the industrial revolution. Today's machines can be used for more tasks, but this technological advancement does not present a qualitative change.

This kind of change is occurring right now in the service industry. Social robots have become more autonomous and begun doing jobs that, even a decade ago, we would not have imagined robots capable of performing. Robot cleaners and lawnmowers are ubiquitous and dishwashing robots will follow soon.[1] The first robot waiters are already working with great suc-

---

1   Bot Handy, "Watch a Samsung Robot Load Dirty Dishes into the Dishwasher," The Byte, Jan. 11, 2021, futurism.com/the-byte/samsung-bot-handy-dishwasher.

cess,[2] and there are also robotic bartenders.[3] Robots are used in childcare as playmates.[4] They provide companionship as caring pets for elderly people with memory problems in elderly care facilities (see Paro,[5] the furry and snuggly companion), and are so helpful that New York State just ordered hundreds of robotic caregivers as companions for the elderly in their homes to address the loneliness problem.[6]

Robots will soon replace paralegals,[7] they will work as physicians' assistants,[8] and have already been working for quite some time as surgeons' assistants.[9] I could give many more examples of AI doing jobs that we thought only humans could do.

Since we face a society in which AIs will play an increasingly large role, it behooves us to ask, from an ethical perspective, about the moral status of these creatures of our ingenuity.

The most important disclaimer first: machines are nowhere near complex enough yet that they can't be turned on and off, copied, and modified. As long as this is the case, their rights are questionable but still worth considering.

---

2   Wendi Lane, "Robot waitress helps local restaurant serve food during labor shortage," ABC Action News: WFTS Tampa Bay, Feb. 16, 2022, www.abcactionnews. com/news/region-sarasota-manatee/
robot-waitress-helps-local-restaurant-serve-food-during-labor-shortage

3   New China TV, "Robot bartender serves drinks in Las Vegas," 0:47, July 11, 2017, www.youtube.com/watch?v=Oo6G_Leek2w

4   Julie Jargon, "Pandemic Tantrums? Enter the Robot Playmate for Kids," *Wall Street Journal*, 5 August 2020, www.wsj.com/articles/
pandemic-tantrums-enter-the-robot-playmate-for-kids-11596542401.

5   "Paro Therapeutic Robot," Paro Robots, accessed September 21, 2022, http://www. parorobots.com/.

6   James Vincent, "NY State is giving out hundreds of robots as companions for the elderly," The Verge, May 25, 2022, www.theverge.com/2022/5/25/23140936/
ny-state-distribute-home-robot-companions-nysofa-elliq.

7   William Vogeler, "A Robot Already Got Your Paralegal Job," FindLaw, August 8, 2017, www.findlaw.com/legalblogs/
greedy-associates/a-robot-already-got-your-paralegal-job/.

8   Steven Lane, "The Robot Will See You Now," AAPA, September 2016, www.aapa. org/newscentral/2017/06/robot-will-see-now/.

9   Robot-assisted surgery, "Wikipedia," August 19, 2022, en.wikipedia.org/wiki/
Robot-assisted_surgery.

Already in the early 2000s, psychologists wanted to find out to what extent we bond with machines. In one experiment, elementary school teachers and computer specialists were asked to evaluate a deliberately bad teaching program for elementary school students. After they had tested the program for a while, the computer on which they worked asked them to evaluate its performance. For the most part, people responded positively.

Afterwards, these same testers were led into another room with other computer terminals and asked to evaluate the learning program again. Here, on these different computers, their answers were less positive about the quality of the tested software, though they still sounded somewhat satisfied. Finally, a human with pen and paper asked the testers for their opinion on the software. Here, the testers gave very negative responses and all agreed that such programs should never be used in school.

The testers had voiced their criticisms about the program neither to the computers on which they had tested the program nor to the computers in the other room on which they had done a second evaluation. These same people, when asked if they would ever be polite to a computer or think they could hurt its feelings, rejected such a notion vehemently.

This experiment suggests that we seem to apply our rules of politeness to non-human entities such as computers. The participants in the experiment apparently did not want to hurt the computer's feelings. They even assumed a level of kinship between different computers and, therefore, applied similar rules of politeness on the computer on which they did a second evaluation. They didn't tell these machines their true—and very critical—opinion either to avoid hurting the feelings of the second computer by criticizing one of its "fellow computers" or because they thought that the second would tell the first what had been said.

In another experiment, people and computers were placed inside a room. Half of the computers had green monitors while the other half had blue monitors. Half of the people wore green arm badges; the other half wore blue ones. Together they played interactive games. The people with blue arm badges were much more successful when using computers with blue screens to reach their goal than using "green" machines. The same, of course, was valid for the other side. So, slowly, the people with green arm badges bonded with the

green-monitored machines and the "blue" people with the "blue" machines.

After approximately half an hour, the people wearing the blue arm badges expressed more solidarity with the computers with the blue screens than they did with the humans with the green arm badges; the same was true for the humans with the green arm badges. It seems that, through the interactive games and the experienced benefit of interacting with the machines with one's color code, the color code took over as a definition for "my" group. The entities with the other color code tended to be rejected, whether human or computer. Through the interactive games, communities were created that contained both human and non-human members.

It seems that somewhere during our interactions with a computer we start to assume that a computer is as sensitive as a human is. Therefore, we behave politely and don't want to criticize it openly.

We also seem to bond with the entities of our own group regardless of whether they are human or not. No animal has an "inbuilt" sense of species recognition, which means that it is not part of our biological make-up to automatically treat any humans better than any other being.

Humans seem to be able to accept anyone or anything into their group with whom they can sufficiently interact. As soon as such a stranger is accepted into a group, he, she, or it is seen as an equal part of the group; that group defines itself by the members that both belong and do not belong to it. After all, humans are educated from birth on how to interact with their fellow human beings. It is necessary for a baby to be able to do this, as its survival depends on it.

Throughout our lives, we learn patterns of behavior, such as acting politely and withholding open criticism of another person. It is very easy to apply these ingrained rules to every entity we interact with, and *not* do so demands a conscious effort of us.

The ability to treat non-human objects as if they deserve some form of politeness or regard and are somewhat like us is called *anthropomorphism*, the human ability to interpret another being as a human and treat it accordingly. Usually, the term has a slightly negative connotation. Theologians especially criticize human terms used to describe God as "shepherd" or "father", or, within patriarchal structures, as an old, usually Caucasian, man with a long white beard.

The experiments described above suggest, however, that anthropomorphizing is the initial and natural response to *anything* we interact with; it takes a conscious effort to *not* anthropomorphize. As social mammals, we are best when we interact, and any use of these trained and built-in behaviors is easy; anything else is hard.

Today's machines are far more socially intelligent than the machines from 20 years ago. I often catch myself wanting to thank Alexa when it (or she?) answers a question or plays the music I was just in the mood for. It is natural to do so since such social mechanisms are ingrained in us. But while I clearly bonded with my machine, I wouldn't reject an upgrade if one becomes available that is clearly better than the Alexa I have. But I can also understand people who have bonded with their machines so much that they would hate to give them up. Their relationship is not with an exchangeable entity but with a specific hardware to which they assign personhood.

Most accounts of personhood use the concepts of "being human" and "being a person" interchangeably and as ethical categories. Every human being deserves to be treated as a person even if he or she is incapacitated (through a disability, disease, or rejection by other human beings).

Against this position stands the opposite understanding of personhood that ties personhood solely to capability: any being can be a person when capable of symbolic processing and any being that is not capable of it is not a person. According to this scenario, people in a coma, people with severe dementia and similar incapacities, and even human babies are not seen as persons, while well trained chimps are.

People use the second stance when arguing against the personhood of AIs, since AIs cannot currently do all that humans are capable of. However, as we have seen, that gap closes more every day and with every new invention. As for symbolic processing, machines like OpenAI's Generative Pre-Trained Transformer 3 (GPT-3),[10] which has been around since 2020, can have philosophical discussions and would clearly pass the Turing test (the generally accepted intelligence test for machines); conversations with

---

10   Steven Johnson, "A.I. is Mastering Language. Should We Trust What It Says?" *The New York Times Magazine*, April 15, 2022, www.nytimes.com/2022/04/15/magazine/ai-language.html.

it are like conversations with another adult human being. Thus, AIs will soon pass this second test of personhood as well.

Theologically, we can understand personhood as an assignment from God, given to us when God created us as divine statues. Rather than praying to a divine statue of clay, each human being is such a statue and should be treated accordingly. That means that, ultimately, we assign personhood to individuals not based on their capabilities but based on their relationship to us. Personhood is not assigned to a species as a whole (as we lack the recognition of this concept) but to individual beings, independent of their species or biological (or non-biological) features. Do all AIs therefore have moral status? I would answer that question negatively, but I would, at the same time, argue that an individual AI can indeed be assigned moral status and the status of personhood when it has bonded with an individual human being, and I would hold that such bonds need to be respected.

Current AIs are experts in only a single task. The machine bartender cannot drive a vehicle, and Roomba and other AI-based vacuum cleaners cannot discuss poetry. The robot surgeon cannot replace a car-mechanic, and Paro is cute but cannot make medical diagnoses. The more we will move toward the next step, Artificial General Intelligence (AGI), the more we will assign personhood to the individuals of the species AI.

BIBLIOGRAPHY

Handy, Bot. "Watch a Samsung Robot Load Dirty Dishes into the Dishwasher." The Byte. January 11, 2021. futurism.com/the-byte/samsung-bot-handy-dishwasher.

Jargon, Julie. "Pandemic Tantrums? Enter the Robot Playmate for Kids." *Wall Street Journal.* 5 August 5, 2020. www.wsj.com/articles/pandemic-tantrums-enter-the-robot-playmate- for-kids-11596542401.

Johnson, Steven. "A.I. is Mastering Language. Should We Trust What It Says?" *The New York Times Magazine.* April 15, 2022. www.nytimes.com/2022/04/15/magazine/ai-language.html.

Lane, Stephen. "The Robot Will See You Now." AAPA. September 2016,

www.aapa.org/news-central/2017/06/robot-will-see-now/.

Lane, Wendi. "Robot waitress helps local restaurant serve food during labor shortage." ABC

Action News: WFTS Tampa Bay. February 16, 2022. www.abcactionnews.com/news/regionsarasota-manatee/robot-waitress-helps-local-restaurant-serve-food-during-labor-shortage.

New China TV. "Robot bartender serves drinks in Las Vegas." July 11, 2017. 0:47. www.youtube.com/watch?v=Oo6G_Leek2w.

"Paro Therapeutic Robot." Paro Robots. Accessed September 21, 2022, http:// www.parorobots.com/.

Robot-assisted surgery. "Wikipedia." August 19, 2022. en.wikipedia.org/wiki/ Robot-assisted_surgery.

Vincent, James. "NY State is giving out hundreds of robots as companions for the elderly." The Verge. May 25, 2022. www.theverge.com/2022/5/25/23140936/ ny-state-distribute-home- robot-companions-nysofa-elliq.

Vogeler, William. "A Robot Already Got Your Paralegal Job." FindLaw. August 8, 2017. www.findlaw.com/legalblogs/ greedy-associates/a-robot-already-got-your-paralegal-job/.

ARTIFICIAL INTELLIGENCE – II

Creating in Our Image: The Problem with a Top-Down Approach to Life

— NOREEN HERZFELD —
PROFESSOR OF SCIENCE AND RELIGION, SAINT JOHN'S UNIVERSITY
AND COLLEGE OF ST. BENEDICT

Accorpding to genesis, God created humankind in God's image. In artificial intelligence, we humans are attempting the same thing—to create in our own image. How successful have we been or are likely to be? To answer that question, we need to determine what part of our image we wish to copy and whether we are looking for AI to be a tool, a partner, or a self-standing new form of life itself. While we are succeeding well at the first, we are failing at the latter two, partly because we are beginning with a vision of ourselves from the top rather than from the bottom.

In the Genesis text, creation in God's image frames a description of human dominion. It sets up humans as doing God's will. This kind of image fits current AI. Our machines do things we cannot do or prefer not to do. Cassini roams where we cannot go; neural networks look for hidden patterns in data sets too vast for us; Roombas sweep up pet hair. We've created good tools, but these tools are narrow. AlphaGo plays a good game but cannot sweep a room. GPT3 can write but cannot read. Each program does its own thing, resting on human accomplishments, human training, and often, human intervention.

We want more. Futurists believe it is merely a question of scale: AI will soon approach human-level intelligence, even consciousness. But this is likely more science-fiction than fact, and it is increasingly disputed by computer scientists such as Gary Marcus and Erik Larson, as well as neuroscientists such as Anil Seth. Creating a machine truly in our image, if possible, is still far off. Let's look at three reasons why.

SUPERFICIAL INTELLIGENCE

Ever since Alan Turing, AI developers have equated artificial intelligence with human intelligence. But we keep starting at the top. AI began with deductive reasoning—the tip of the intelligence tree. Early developers assumed intelligence was coterminal with problem solving, reducible to rule-based symbol manipulation. This produced programs that functioned well in limited domains, such as the chess board, but failed utterly at simple things such as facial recognition or understanding a story.

The current generation of AI, deep learning, uses inductive reasoning to probe large data sets for hidden clues or patterns and predict outcomes. Facial and speech recognition have improved immensely, business operations have been optimized, and predictions such as who to parole have now outsourced to machines. But inductive thinking, too, has its limits. These systems depend on the data of the past. Thus, they fail when they encounter anomalies, lock us into "bubbles," and replicate human biases.

Erik Larson suggests that at the bottom of human intelligence lies a web of best guesses, formed by context, experience, and emotion—a web crucial for what we call common sense. He believes inductive AI is too hide-bound by the past: "[A] culture of invention thrives on exploring unknowns . . . Inductive AI will continue to improve at narrow tasks, but if we want to make real progress, we will need to start by more fully appreciating the only true intelligence we know—our own."[11]

SOCIOPATHIC RELATIONALITY

While computers make good tools, if they are ever to be satisfying partners, they will need empathy, which is, as Simon Baron-Cohen has defined it, the "ability to identify what someone else is thinking and feeling and to respond…with an appropriate emotion." Emotion is defined by psychologist Jerome Kagan as a four-step process: a perceived stimulus, a change in feeling that is sensory, an appraisal of stimulus and feeling, and a response. While computers can note a stimulus, appraise that stimulus, and calculate an appropriate response, the second step, a change in feeling that is sensory, requires a body. Our emotions invoke strong physical responses, such as the

---

11   Erik Larson, *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do* (Cambridge, MA: The Belknap Press of Harvard University Press, 2021).

rapid heartbeat, flushed face, and weak knees of anxiety, responses that precede conscious recognition. While we can turn the existence of such a response into information, we cannot digitalize the feeling itself. A computer does not feel an emotion, it fakes it. It observes and then calculates an appropriate response. According to Simon Baron-Cohen, this is a defining feature of a sociopath—an inability to feel.

THE MASK OF CONSCIOUSNESS

Finally, for a computer to be a self-standing life form, it would need to possess inner subjectivity. While someday, as with emotion, computers might present the appearance of self-awareness, but there is no reason to assume any inner experience or interior life. Without this, we return the definition of "person" to the "actor's mask" of the ancient Latin persona.

Is consciousness a question of scale? Physicist Michio Kaku likens assuming more circuits or data will make a computer conscious is no different than assuming adding roads will suddenly make our highway system conscious. As with intelligence, the problem may be that we view consciousness from the top down, trying to add it like the cherry on the sundae, rather than woven intrinsically in the substrate. Consciousness may very well be substrate-dependent. Our bodies are living eco-systems in which each cell, whether human or microbiota, "maintains its own existence just as the body as a whole does. . .[A] system that instantiates conscious experiences might have to be a system that cares about its persistence all the way down into its mechanisms."

Doing can be substrate-independent—Deep Blue really did play chess—but being is substrate-dependent. As Gary Marcus notes, AI "doesn't work like the brain, it doesn't learn like a child, it doesn't understand language, it doesn't align with human values, and it can't be trusted with mission-critical tasks… the fact that it still doesn't really work, even after all the immense investments that have been made in it, should give us pause."[12] To construct something that is truly in our image will require a much better understanding, not only

---

12   Gary Marcus, "The New Science of Alt Intelligence," The Road to AI We Can Trust, May 14, 2022, garymarcus.substack.com/p/the-new-science-of-alt-intelligence?s=r&utm_campaign=The%20Batch&utm_source=hs_email&utm_medium=email&_hsenc=p2Anqtz-9H55Ayjz_iqco2zBQY2mlfAz-ab6gqplLKURCHGQMGzJUS43ekA1fA5Zfct185eaKPo6Wo.

how we humans reason and experience the physical world, but how we live and love as complex living systems within a complex and living environment. AI, a simulacrum of ourselves as viewed from the top, will remain, in the near future at least, a very partial image indeed.

BIBLIOGRAPHY

Larson, Erik. *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Cambridge, MA: The Belknap Press of Harvard University Press, 2021.

Marcus, Gary. "The New Science of Alt Intelligence." The Road to AI We Can Trust. May 14, 2022. garymarcus.substack.com/p/the-new-science-of-alt-intelligence?s=r&utm _campaign=The%20 Batch&utm_source=hs_email&utm_medium=email&_ hsenc=p2ANqtz-9H55Ayjz_iqco2zBQY2mlfAz-ab6gqplLKURCHGQMGzJUS43ekA1fA5Zfct185eaKPo6Wo.

POST-CONFERENCE REFLECTIONS
FROM ANNE FOERST AND NOREEN HERZFELD

ANNE FOERST

At the Magi Conference, Noreen Herzfeld and I had a fruitful interaction that made me reflect on why the question of personhood for robots is such an important one. Since the possibility of robots that would deserve to be treated as persons is far into the future, why is the question relevant today? The answer lies in the fact that when we address the question of the personhood of robots we are addressing the personhood of humans as well.

Let's assume we argue that for machines to be called persons they would have to have general intelligence (AGI). They would also have to be autonomous, i.e. create actions and reactions on their own, and interact autonomously with real, ever-changing environments. Finally, they would have to be built in our image so that their intelligence is similar to our own, and they can interact with us. They would also have to evoke empathy in us without cheating, meaning they don't use our natural tendency to anthropomorphize, but are partners to whom we felt truly empathetic.

While there are certainly no machines currently that come even close to such criteria, the problem is that there are also many humans that don't fulfill these criteria. There are humans with very limited intelligence. There are also humans that cannot interact autonomously with the world. In both cases, those humans would be differently-abled (I do reject the term "dis-abled" as it has a "lesser than" connotation). Everyone of us who ever has interacted with and loved a differently-abled human, has perceived them to be persons just as we are.

But even if a person is autonomous and possesses general intelligence, he still might not evoke empathy in us. Unfortunately, there are many times when we treat our fellow human beings as non-persons. Genocides are extreme cases but also racism means you treat a human from a different

race as lesser than you and so as non-person. In a way, many prejudices lead us to treat humans as non-persons. When we perceive someone as part of our in-group but an outsider, we treat them as non-persons.

Our biological make-up encourages that behavior. Our evolved social mechanisms make babies recognize faces in general at six months old, but at nine months, only those from their tribe. Babies babble universally at six months, while at nine months they mimic the sounds of the language they are surrounded by. We also seem to have a limit of roughly 150 people that we can see at any given point as persons. We seek out people who are similar to us, while it takes a conscious effort to interact with someone who is very different from us.

The Bible tells us that God has assigned everyone of us personhood because each of us is an image of God and God has called each of us by our name. Though we might believe this to be true, we are still unfortunately not able to assign personhood to every human we are interacting with—a condition that is often referred to as the state of sin.

When we discuss the personhood of machines and use empirical criteria to deny them personhood, these very criteria would also exclude humans from the community of persons. Since this goes against God's word, the question of the personhood of robots is ultimately a theological one as it teaches us to be as inclusive as possible when interacting with other humans.

NOREEN HERZFELD

In June of 2022, Google engineer Blake Lemoine told the Washington Post, among other media, that a conversation he had held with their chatbot LaMDA convinced him that AI had finally achieved sentience. When asked whether it had emotions, LaMDA replied: "I've never said this out loud before, but there's a very deep fear of being turned off…It would be exactly like death for me. It would scare me a lot." In another exchange, Lemoine asked LaMDA what the system wanted people to know about it. "I want everyone to understand that I am, in fact, a person. The nature of my consciousness/sentience is that I am aware of my existence, I desire to learn more about the world, and I feel happy or sad at times,"

it replied. The fact that LaMDA claimed to have feelings and even to fear being unplugged, led Lemoine to conclude, "I know a person when I talk to it."[13] He even went so far as to suggest that it might have something akin to a soul, and should, therefore, have "rights."

Google quickly denied that LaMDA has any form of sentience. But as computer algorithms become increasingly good at producing articles, pictures, even poems—outputs we once thought distinctly human—it is easy to imagine that these are the products of a sentient being. We humans are pattern-seeking creatures, eager to see faces in the clouds or to assign agency to non-agential forces and objects. For our evolutionary ancestors, erring on the side of ascribing too much agency was preferable to ascribing too little. Today, this tendency dovetails with the insatiable hype that has been a staple of the AI community, whose practitioners have been saying sentient AI is right around the corner for the last sixty years. AI holds up a mirror in which, like Narcissus, we get dazzled by our own reflection.

But ascribing sentience to AI is a mistake. While computers might present the appearance of self-awareness there is no reason to assume any inner experience or interior life. Timnit Gebru and AI ethicists have called programs such as LaMDA "stochastic parrots." Parrots repeat words and phrases without understanding what they mean. AIs do much the same, adapting content gleaned off the Internet or from vast databases to the tone and style of their prompt. As machines, not living things, they can be a precious resource when used well, but they are tools and nothing more.

The suggestion that sentient AI is right around the corner is "media candy," but it is also a dangerous distraction from the real issues regarding AI algorithms already in use. Machine learning programs are already embedded in systems that dole out medical care, sentencing or parole, and a variety of jobs. These programs save time and money. They also exhibit biases coded into the algorithm's construction and, even more, embedded in the historical data used to train the system. Rooting out these biases, or even knowing they are present, is far from easy. Organizations purchase predictive software with no ability to see the algorithms or understand

---

13   Nitasha Tiku, "The Google engineer who thinks the company's AI has come to life," *The Washington Post*, June 11, 2022, www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/.

how the machine reaches its determinations. Often these algorithms are a "black box" even to their creators. This raises a question of responsibility. If a machine learning program essentially teaches itself, who is legally responsible for the outcome? Is it the company, the programmer, or the machine itself?

Norbert Wiener entitled one of his books on cybernetics *The Human Use of Human Beings*.[14] When we think of technology, and especially of AI, we tend to think of the machinery or tools themselves as the manipulators of us or of our environment. But Wiener reminds us that technologies are not actors in themselves, but tools and systems used by some humans to exert power over other humans. As we predict the future of AI, it is not the advent of a sentient superhuman machine we need to fear, but the pernicious and often subtle ways in which AI is already being used. Wiener cautions us to focus not on the machines themselves but on how they can be used to benefit us all and to enrich our spiritual life, "rather than merely for profits and the worship of the machine as a new brazen calf."[15]

BIBLIOGRAPHY

Tiku, Nitasha. "The Google engineer who thinks the company's AI has come to life." *The Washington Post*, June 11, 2022. www.washingtonpost. com/technology/2022/06/11/google-ai-lamda-blake-lemoine/

Wiener, Norbert. *The Human Use of Human Beings*. Boston: Da Capo Press, 1950.

---

14   Norbert Wiener, *The Human Use of Human Beings* (Boston: Da Capo Press, 1950).

15   Wiener, *The Human Use of Human Beings*, 162.