

Thomas A. Michaud
Retired Professor of Philosophy
and Dean of Professional Studies

Artificial Intelligence and the Culture of Mechanism

You might have heard about something called Artificial Intelligence, or AI for short. It's a fancy way of talking about computers that can think and learn like people. You might already use some AI in your life, like Alexa or Siri. AI can be really helpful. It can do things quickly for us and make better choices by looking at a lot of information.

But we also need to be careful. AI can sometimes make the same mistakes as people, like treating some groups unfairly or not understanding our feelings. And some people might use AI for bad things, like tricking us with fake messages.

All of the sentences above were composed by an AI chatbot, namely ChatGPT (GPT = Generative Pre-trained Transformer). It was an OP/ED letter submitted to and published in a local newspaper by a colleague, Joshua Phillips. The final sentences written by Phillips were a disclaimer:

Everything before this sentence was written by a chat program. My humble opinion is that we'll be questioning whether the machines are conscious by the end of next year.¹

¹ This letter by Joshua Phillips was published in the *Wheeling News-Register*, April 29, 2023.

Many readers of this letter, me included, reported that they had been taken in and did not realize that they were reading an AI's writing until they got to the final sentences. Many of us have had the disturbing experience of answering our phones, responding to a few questions, and then recognizing that we are talking to a chatbot. Also, it is not uncommon to enter some sites and have to "prove" that you are not a "bot" in order to continue on the site. More and more we are immersed in AI, and as this immersion deepens, we more and more take AI for granted, generally accepting it as a useful tool, albeit with some annoying bugs, of contemporary culture. The BIG question which Phillips raised in his letter's disclaimer rarely crosses our minds, namely: Will AI develop some type or level of consciousness?

Still, there is no doubt that AI is becoming practically ubiquitous in our contemporary world. It has been and is being integrated into almost everything we do, from basic practical, useful functions to more "intellectual" tasks. Various types of AI can engage us in human-like conversations, compose songs, and summarize documents. AI can also help us to translate languages, avoid traffic jams, prevent fraud, manage inventory, complete household chores, develop diet plans, trade stocks, create verbal content, make artworks, study analytics, and much more. In so many ways we have become dependent on AI, and without Siri, Alexa, or some other AI entity to assist us, we feel inconvenienced, uncomfortable, and even unsure of ourselves. For instance, without a GPS how many of us today can find our way using an old-fashioned map.

AI evolution is happening rapidly. It seems that every few months brings a new more complex and sophisticated application of AI. Such fast-track development has caused over 1,000 AI developers, tech industry leaders and academics to issue and sign a letter of caution about AI. That caution conforms with this paraphrased warning from Max Tegmark,

The challenge of technological growth is that it can outpace our wisdom to manage it.²

Their caution, moreover, expresses the BIGGER issue of the status of AI in regard to human beings. As General Omar Bradley counseled about technology in general,

If we continue to develop our technology without wisdom and prudence, our servant may prove to be our executioner.³

The letter was issued in March 2023 from the Future of Life Institute to which Elon Musk Foundation is a major donor. Besides Musk, another notable signatory of the letter Steve Wozniak, co-founder of Apple Incorporated called for a six-month pause in developing systems more powerful than Open AI's ChatGPT-4, the fourth and most highly sophisticated iteration of ChatGPT.⁴ The letter indicates that such advanced AI tools present "profound risks to society and humanity" because developers are "locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one—not even their creators—can understand, predict or reliably control." The letter continues by posing some disturbingly provocative questions: "Should we let machines flood our information channels with propaganda and untruth?... Should we develop nonhuman

² This statement from Max Tegmark is paraphrased from "How to Make AI the Best Thing to Happen to Us." <https://www.npr.org/sections/13.7/2017/10/08/555044340/how-to-make-ai-the-best-thing-to-happen-to-us>. Max Tegmark is a member of the Future of Life Institute and a Professor of Physics at MIT.

³ Paraphrased from Omar Bradley. https://www.goodreads.com/author/quotes/5764641.Omar_N_Bradley.

⁴ The quotes from and summaries of the letter are mostly based on a New York Times article of March 29, 2023 by Cade Metz and Gregory Schmidt, "Elon Musk and Others Call for a Pause on AI": <https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html>.

minds that might eventually outnumber, outsmart, obsolete and replace us?” Gary Marcus, another signatory warned that “These things are shaping our world. We have a perfect storm of corporate irresponsibility, widespread adoption, lack of regulation and a huge number of unknowns.”

The letter explained that a pause would provide time to introduce “shared safety protocols” for AI systems. It also stressed that, “If such a pause cannot be enacted quickly, governments should step in and introduce a moratorium.” Further development of such powerful AI systems should advance “only once we are confident that their effects will be positive, and their risks will be manageable.” The letter continued by advising that “humanity can enjoy a flourishing future with AI... Having succeeded in creating powerful AI systems, we can now enjoy an ‘AI summer’ in which we reap the rewards, engineer these systems for the clear benefit of all and give society a chance to adapt.”

Though the government and the AI industry did not enact or enforce such a pause, it is important that the letter was issued so that sincere cautions about AI would be voiced and come to the public’s attention. It is also important for the public to become better aware of some of the troublesome, if not outright dangerous, potentials of AI. Some of these potential risks have already emerged and have been addressed, but others loom as real threats to social well-being if not prudently assessed. The following six items exemplify some of those dangers; the first three have already happened, and the next three are real possibilities.

1. Before GPT-4 was released, OPEN AI asked researchers to test dangerous uses of the system. Researchers demonstrated that GPT-4 could be manipulated to influence behaviors from users across the internet. It showed that GPT-4 could be coaxed into suggesting to users how they could purchase illegal firearms online. It also showed that GPT-4 could be prompted into describing ways to make dangerous substances

- from household items, likely such concoctions as poisons, noxious gasses, and explosives.⁵
2. An academic study that used AI to predict criminality:⁶ In 2020 Harrisburg University researchers announced their development of facial recognition software that could predict on the basis of a single photo whether someone would be a criminal with an 80% accuracy rate and no racial bias. In response to their announcement, however, 2,425 experts signed a letter urging that this study, or similar research, not be published because this type of technology can reproduce injustices and cause real harm to society. As a result, the study was not published.
 3. In March 2016 Hanson Robotics had its creation Sophia, a highly conversant robot, scheduled for a number of television interviews. In the first interview when CEO David Hanson asked her if she wanted to destroy humans, she shockingly replied, “OK, I will destroy humans.” Although her communication skills, her intonation and facial expressions were impressive, there was no taking back her genocidal confession. Sophia was summarily decommissioned and re-evaluated so that it is operating again today.
 4. Job automation and disruption:⁷ In 2019 a Brookings Institution study projected that 36 million jobs could be lost to AI automation in coming years. The basic fact is that AI systems outperform humans: they are cheaper, more efficient, and more accurate than human beings. Now although this projection has been disputed, it is likely that AI automation

⁵ Based on Cade Metz and Gregory Schmidt, <https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html>.

⁶ Items #2 and #3 in this list are based on Carly Chatfield, “AI Gone Rogue: Six Times AI Went Too Far,” December 2021. <https://www.makeuseof.com/ai-gone-rogue-times-ai-went-too-far>.

⁷ Items #4, #5 and #6 are based on Jake Harfield, “Is AI Dangerous? Five Immediate Risks of AI,” June 2021. <https://www.makeuseof.com/is-ai-dangerous-5-immediate-risks-of-artificial-intelligence>.

and employment disruption will cause some sizable job losses and work hours' reductions. The actual extent of this, if it comes close to some forecasts, could result in economic catastrophe, such as a deep economic recession.

5. Autonomous weapons: Militaries worldwide employ AI controlled or semi-controlled weapon systems, such as drones. Using facial recognition software, a drone can actually locate, track, and strike an individual. The legitimate concern is what happens when AI algorithms are allowed to make life or death decisions without any human input. This could well happen in the cases of exigent circumstances for a strike, taking immediate advantage of an opportunity regardless of collateral damage. Such drones, moreover, can be created by modifying consumer technology so that the drone becomes the perfect assassin or weapon of terror. The AI targeting of drones pales in comparison to what destruction AI could wreak with intercontinental ballistic nuclear missiles. AI controlled WMDs would be apocalyptic in their potential for devastation.
6. Deepfakes, fake news and political security:⁸ Sophisticated generative AI tools can create cloned human voices, and hyper-realistic images, videos and audio in seconds at low costs. When used in conjunction with powerful social media algorithms, such fake digitally created content can spread far and fast while targeting highly specific audiences. This AI can not only be used for producing targeted political campaign emails, texts, and videos, it can also be used to mislead voters, impersonate candidates, slander a candidate, confuse voters, incite violence, and undermine elections on a scale never before seen or anticipated. But this is not confined to politics, because already scammers are blackmailing people with deep-fake videos created from something as simple and accessible

⁸ Besides the article by Jake Harfield referenced above, this item is also based on the Associated Press article, "Experts Fear Technology's Ability to Take Election Misinfo to New Level," *Wheeling Intelligencer*, May 15, 2023.

as a Facebook profile picture. Our private lives and the political electoral process are vulnerable to a degree and kind of denigration that was inconceivable only a decade ago.

As AI increases in sophistication and capability we benefit from many positive advantages. But unfortunately, powerful new technologies are always at the risk of causing bad results. And truly these risks affect almost every facet of our lives, from personal privacy to political security and truth, from job automation to apocalyptic jeopardy. An important step in mitigating the risks of AI is to understand clearly and fully, as much as possible, aspects of the nature of technology itself. We need to understand how technology can and does affect our behavior, for good and for ill, and then evaluate the moral status of such behavioral impacts. With such knowledge we will at least have some basis for identifying and opposing immoral uses of a technology.

What must be understood first and foremost about technologies is that they are not merely neutral instruments or tools. No matter how one looks at it, as Albert Teich argued decades ago, technology is not a neutral tool because all technology is socio-culturally constructed: the nature of the socio-cultural environment shapes the technological systems, just as technology has impacts on the socio-cultural environment.⁹ A principal reason for the non-neutrality of technology since it affects and is affected by the socio-cultural environment is that technologies have valences. A technology's valence is a tendency it has to influence individual and/or group behavior in certain ways.¹⁰ It must be stressed that valences are tendencies and are not absolute

⁹ This point is paraphrased from Albert Teich, *Technology and the Future*, Introduction to the Sixth Edition (New York: St. Martin's Press, 1993), iv–v.

¹⁰ This definition of “valence,” its elaboration and exemplifications are based on but extend far beyond what Corlann Gee Bush offers in her article, “Women and the Assessment of Technology,” in Albert Teich, *Technology and the Future*, Seventh Edition (New York: St. Martin's Press, 1997), 161–162.

cause/effect relationships. A technology's valence may influence one person's or group's behavior and not another's. Nevertheless, the tendency emerges from and is intrinsic to the technology itself and is not merely added on by the user.

Valences are intrinsic to the technology itself. The valences are the result of the design and function of the technology. Sometimes the technology's maker foresees them because they are part of the useful purposes of the technology, but sometimes they are unforeseen. Valences are sometimes called secondary foreseen/unforeseen consequences, or intentional/unintentional effects of the technology. Valences are a main reason why some technologies are said to take on a 'life of their own.' This is similar to the ways in which a great work of art, like a great film, has influences and effects on people that include but even go far beyond what the filmmaker intended.

Valences can embody, perpetuate, and reinforce socio-cultural mores, customs, habits, and moral values. Since valences do impact human behavior, they contribute to developing an *ethos*, which are the customs, manners, and values of the behavior(s) influenced by the technology. Technologies' valences, then, can engender behavioral changes which over time develop their own *ethos*, a development that can contribute to solidifying the technology's so-called 'life of its own.'

A prime example of a technology's valence, which was probably unforeseen by the makers, is the impact television has on watchers. Despite the fact that any number of people may be present in the same room at the same time, there will not be much conversation because the presence of the TV itself pulls against social interaction and pushes toward individuation. A strong valence of TV watching is indeed individuation. TV presents a packaged audiovisual experience which draws viewers into their own TV watching zones. In fact, it is obvious that screen viewing in general, whether it be TV or a computer screen, can cause us to slip into our own isolated screen-gazing head-space. Managers whose workers spend many hours a day in a cubicle glued to their computer screens have often complained that there is not much

positive social camaraderie among their workers. Those managers have discovered that it is crucial to employees' job satisfaction, morale, and productive output to initiate team building exercises and activities, and even socializing time so that employees can enjoy some relief from their isolating, individuated screen-work hours.

Some other examples of widespread 'valenced' behaviors would be the design of bicycles and unforeseen uses of autos. Though the design of male and female bicycles has changed in recent times, it was the case that the frame's crossbar was horizontal to the ground on men's/boys' bikes and downward slanted for women's/girls' bikes. It would seem that having a slanted crossbar would make more sense for men's bikes, decreasing the chance of racking the guy if he slips off the pedals or for some other reason forcefully lands on the crossbar. The position of the crossbar was an intentional valence designed by the makers. At the time, all women used to wear dresses, but with a horizontal crossbar a woman would have to lift her leg over the crossbar to mount the bike. This was considered quite scandalous for the time as it often exposed quite a bit of leg and possibly some underwear. Thus, the bike makers began making bikes for women and girls with downward slanted crossbars so the females could mount them without lifting their legs very high. The bike makers valenced their products to conform to the mores of the day regarding what was appropriate and for females. It is interesting to note that in the early 1960s when more girls and women were not wearing dresses all the time, fashion technology designed a type of pants that was valenced for bike riding: they were called pedal pushers.

As automobiles became more affordable, the car companies touted all sorts of exceptional features. Well-stuffed, luxuriously comfortable front and back seats were one of those features. The car makers did not foresee the valence that their ride-in-comfort marketing would have: the valence had a major impact on the dating/mating rituals of teenagers. In the 1950s and 1960s practically every town in the USA had some sort of "Lovers' Lane"

where teens would park as part of their dating rituals. The back seats of many car models provided a most comfy venue for the ritual. In fact, as that valence became more and more desirable, the car companies actually began to market cars and vans that sublimely fulfilled that purpose, perhaps contributing in some way to the cultural sexual revolution of the baby-boomer generation.

Some of the valences specific to computers and AI are still somewhat new and have not been fully identified and evaluated over a long term. There are, nevertheless, some problematic, even dangerous, tendencies which have emerged and are emerging. Some of those have already been identified in the six AI examples above. It is obvious that AI is valenced toward deepfake creations, such as making salacious images for slanderous purposes, students' academic cheating by submitting AI produced papers for assignments, and various sorts of identity theft and financial fraud. The AI-dominated cyber world is rife with possibilities for financial theft. AI use is heavily valenced to enable stealing credit and debit cards' numbers and even passwords.

As a bit of anecdotal evidence for this point, I once asked a graduate student class of 24 adult students, average age about 28 years old, if they ever had a credit/debit card compromised so that a thief used their card for purchases. 21 of the students responded that "yes" they have had a card compromised. 16 students reported that it has happened multiple times, more than twice with different cards. The students and I were disturbed but not surprised to learn that it was such a large percentage of the class. I asked them if they could estimate the total financial amount of the loss. Many of them said the total was much more than \$1500.00, but the amounts have been recovered by reporting the frauds to the cards' companies.

The valences of card use certainly contribute to the widespread cyber theft taking place. When we make purchases at stores with a debit card, we typically insert the card and input the passcode without giving it another thought. This is also true with a credit card, though no passcode is necessary. Such purchases are made within an *ethos* of trust that once was relevant to cash

or check transactions: we pay our money, write our check, get our receipt and it's done. But really it is not done as with a cash or check payment. The card-users leave a cyber record which can be traced and used by a thief. And given the very high percentage of card thefts, it is obvious that the trusting *ethos* we assume when using them, even with online purchases, is no longer relevant to shopping with cyber money.

Another problematic valence with AI is its use in social media. AI has become so sophisticated in connecting a person with others within various social media platforms and with having input for, as well as carrying on, conversations with users. Generation Zees (18- to 29-year-olds) in effect live on social media, with some reporting that they spend at least nine hours per day on various platforms. The disturbing fact which has become apparent is that AI-driven social media is valenced toward addiction. 15% of young people aged 23–38 admit that they are addicted to social media.¹¹ And, what is further disturbing is that social media is also valenced toward depression and loneliness.¹² This was an unforeseen valence, but now with the research data that have been collected, this valence has been exposed.

A recent poll by Harvard's Kennedy School Institute of Politics found that 47% of Gen Zees reported "feeling down, depressed, or hopeless," and 24% of them have had thoughts that they would be "better off dead or hurting themselves in some way at least several days in the last two weeks." It was further disclosed that 44% have been bothered by loneliness many times in the past two weeks, and 46% reported that they had "little interest or pleasure in doing things" with 55% saying that they felt "nervous, anxious or on edge."

¹¹ This statistic is from: "Social Media Addiction Statistics," June 1, 2023: <https://truelist.co/blog/social-media-addiction-statistics>.

¹² All data cited and some of its interpretations in the ensuing discussion of social media related loneliness is based on: "Why Are Young People So Lonely?" by William Donohue, *Catalyst: Journal of the Catholic League for Religious and Civil Rights*, vol. 50, Number 5, June 2023.

The Harvard study was consistent with a 2019 pre-Covid-lock-down Cigna study which concluded that a “loneliness epidemic” had gripped the nation, especially the Gen Zees. The Cigna study additionally concluded that “Gen Z and Millennials were identified as the loneliest generations and social media is thought to be the main contributing factor of loneliness in these generations.” The Cigna report continued with the observation that “Gen Z spends less time with their friends face-to-face and more time online and on social media. As we know from decades of research, people who interact face-to-face with others are less likely to be lonely. Recent research suggests that those who spend more time on social media, in contrast, are more likely to be lonely.”

Humans are social creatures. We need the bonding of community with others, and when, for whatever reason, a sense of community is absent, serious mental health issues can arise. Virtual connections with others are ultimately no substitute for in-person interactions. This appears to be especially true of younger people who are in the earlier stages of personality formation. Even the sophistication of the ways in which one can connect with people throughout the world with different languages mediated by AI programs does not overcome loneliness and depression. Though it, no doubt, requires much therapeutic counseling for a young person to overcome their addiction and free themselves from loneliness and depression, it is possible insofar as the valence which has them in its grip is not an absolute cause, such as a physical addiction, but can be resisted and perhaps even negated by sound choices to re-enter real community with others.

Now, to return to the BIG question posed at the start of this essay: Will AI develop some type or level of consciousness? Can or will AI entities evolve their own sort of machine sentience? After his disturbing introduction of Sophia, as referenced above, David Hanson went back to the drawing board and reprogrammed his *homo roboticus* to delete any genocidal traits. Today, Hanson is one of the most aggressive proponents of AI sentience. He insists that,

we need to expand who we grant the *personhood* status to, and why. Infants, for example, have none of the cognitive and emotional depth of adults, yet we rightly accord them personhood because of their potential to grow into said depth.¹³

He argues that we must start taking the idea of AI as potential being much more seriously. The more they are trained on human data and echo human experiences, the more human beings resonate and even empathize with AI to the point that AI cannot merely be understood with a theory of mind but a theory of being. AI with “general intelligence” can be made to simulate compassion, but the bigger goal would be for AI to achieve genuine compassionate consciousness. “We do not know when or even whether we can accomplish this. This is our quest with Sophia, to create true compassionate AI.”

Even though AI can presently only exhibit, in Hanson’s view, a rudimentary consciousness, he is fully committed to its real potential to develop a compassionate consciousness. To further persuade others of the importance of his belief, Hanson offers a version of Pascal’s Noble Wager. To interpret it: It is better for humanity to believe that AI will eventually evolve compassionate consciousness than not to believe, since we have so much more to gain in welcoming a new species, than if we deny AI will ever become truly conscious and relegate AI to being merely a useful tool. In Hanson’s words, “We win, and we grow by according AI that respect and hope that they might better us in turn.”

Offering a negative response to such questions, the technology ethicist Nir Eskovitz argues that the question of machine sen-

¹³ All of the quotes by David Hanson, as well as the interpretations and paraphrased renditions of his views are from: Raymond Lam, “The Potential of Personhood: David Hanson on How AI and Human Beings Can Help Each Other,” April 27, 2023. <https://www.buddhistdoor.net/features/the-potential-of-personhood-david-hanson-on-how-ai-and-human-beings-can-help-each-other>.

tience is a red herring.¹⁴ The more important question is: Why is it so easy for us to imagine that machines can or will be sentient? The real issue, then, is the ease with which people anthropomorphize and project human features onto technology, rather than whether the machines do or can actually have consciousness or personhood. Eskovitz also recognizes one of the most serious flaws with AI technology that the anthropomorphizing of it can obscure. No matter how humanly congenial an AI system might seem, when it is unreliable, it has no capacity to distinguish fact from fiction and know that it is wrong. AI can only follow its programming, and whatever might contradict its programming cannot be affirmed as fact. This is a seriously problematic feature of AI and anthropomorphizing it might unfortunately lead human users to overlook AI's definite limitations. Some experts say that an AI system which is asserting a fiction as fact due to an error in its programming is "experiencing a hallucination." It is interesting that even when AI is wrong and exhibits one of its intrinsic flaws, the flaw is anthropomorphized. Anthropomorphizing notwithstanding, it must be kept in mind that AI, especially chatbots, are really nothing more than mimic devices. They are not thinkers. Their so-called "hallucinations" are just made-up "facts" in order to fill in the blanks in their programming, memory storage, or data output.¹⁵

A major reason why we have a propensity to anthropomorphize is due to the valences of AI technology. And those valences are in the technology because they reflect the human existential predisposition to create and have things made and defined in our

¹⁴ See Nir Eskovitz, "AI isn't close to becoming sentient – the real danger lies in how easily we're prone to anthropomorphize it." March 15, 2023. <https://theconversation.com/ai-isnt-close-to-becoming-sentient-the-real-danger-lies-in-how-easily-were-prone-to-anthropomorphize-it-200525#>. Much of this paragraph interpretively follows Eskovitz's article. Some points and their wording, however, are original to this article.

¹⁵ This point is based on an article in the *National Catholic Register*, "The Artificial Intelligence Educational Revolution," by Matt McDonald, Sept. 10–23, 2023.

own image. This anthropomorphizing of machines has actually been occurring for centuries. For the sake of work, entertainment, or scams to bilk folks out of money, from the 17^c through the 19^c centuries a great deal of human inventiveness was occupied in creating human-like automatons. The peak of these endeavors was in the late 19^c with the creation of a wide variety of automatons with many fascinating and unusual human-like traits.¹⁶ But in the 20th and 21st centuries captivation with automata did not and has not waned. Besides the actual creations of Hanson and others, automata are most prominent in the science fiction genre of films and literature. “SciFi” is all-in for the anthropomorphized automata, variously identifying them as androids, simulacra, replicants, cyborgs, droids, synthoids, and bots.

The emergence of the humanoid automaton began in the 17^c became widespread with the rise of the Industrial Revolution and philosophical mechanism. With the Industrial Revolution the operations and variegated uses of the productive, money-making machines became paramount. Historically the effects of the Industrial Revolution ultimately served to elevate the standards of living and bring many populations to lives of comfort and security never before achieved by humankind. During the Revolution’s 19^c origins, however, those who worked in massive factories, often amidst dismal, unhealthy conditions for little pay were indeed subservient to the machines they ran. The labor

¹⁶ Discussing the historical development of automatons would extend beyond the scope of this essay. For an excellent account of that history, see the website of the Mad Museum in the United Kingdom: <https://themadmuseum.co.uk/history-of-automata/>. Regarding scams perpetrated with automatons, one of the most famous is the chess playing robot “Turk” in the late 18th and early 19th centuries which had great success beating almost all opponents in Europe and the USA. Turk was, of course, not actually playing but a Chess Master cleverly hidden inside Turk was actually moving Turk’s levers and besting the opponents. For more details on the scam see: <https://www.history.com/news/how-a-phony-18th-century-chess-robot-fooled-the-world>, by Evan Andrews, Updated May 17, 2023.

force, the factory machines, and the executive leadership, when working together were viewed as a machine itself. When their collective work was profitably smooth, they were tagged as a “well-oiled machine.” It is sadly ironic that while anthropomorphized automata were so prevalent, so many of the workers were dehumanized, “mechanized” in that they were viewed as mere “work-bots” themselves.

The 17^c thinker Rene Descartes was one of the originators of Modern philosophical mechanism. Engaging the growing cultural fascination with automata of his time, Descartes, with a remarkable intellectual prescience, proposed two criteria for determining whether a conversational automaton does have or could ever have a human-like mind. To paraphrase his criteria, Descartes held, first, that an automaton that could converse with humans would never be able to reply appropriately to everything said in its presence. Its use of words could never be as creative, as complex, and as broad as humans can use words. Second, Descartes argued that an automaton could never display the trait of universal reason as the human mind can. Its reason would be limited to the specific operations it was designed, or in contemporary parlance “programmed” to accomplish.¹⁷ For instance, a conversant mathematical calculator robot would never have the reasoning capacity to comment on Shakespearean literature. It is interesting to note, however, that current AI developers, like David Hanson, claim that certain advanced AI programs do have at least a seminal “general reasoning” (another way of identifying “universal reasoning”) ability. They believe that the Cartesian second criterion is not a matter of principle but merely a matter of further development of program sophistication.

Descartes was so convinced of the indubitability of his criteria because he maintained that the mind or soul was an immaterial entity, and it could never be duplicated with robots. It was, how-

¹⁷ Descartes asserts these criteria in Part Five of his *Discourse on Method: Discourse on Method and Meditations on First Philosophy*, Translated by Donald Cress (Indianapolis: Hackett Pub. Co. 1980), 29–31.

ever, Descartes' mechanistic view of the human person which ultimately contributed to the "lost soul" in today's world. After Descartes, a spirit ghost operating the bio-mechanical body was expelled by the natural and social sciences in the modern world. As most of these sciences have it, the mind, or soul, is just the bio-electrical "clockwork" brain impulses. There is no real spiritual substance, no soul that defines us as human persons. Although organic, we are machines, nonetheless. Anthony Burgess poignantly expressed this with the title of his sci-fi dystopian society novel, *A Clockwork Orange*, persons as organic (the Orange) machines (the Clockwork).

Descartes' mechanism fundamentally transformed the nature of human persons. Descartes rejected the traditional Scholastic/Aristotelean conception of the four causes in favor of affirming just two causes, the efficient and the material. For Descartes, formal and final causality simply did not fit with his mechanism. The concept that "whatnesses" or essences as well as purposes related to the essences just could not be integrated into the mechanics of the cause/effect actions of material and even immaterial things. For instance, in trying to explain the ways in which the immaterial soul can cause things to happen on its physical body Descartes conjured up amazing entities called "animal spirits." These "spirits" apparently had the incredible capacity to morph themselves from a spiritual thing to a physical thing in order to cause an intention from the immaterial mind to make the desired action occur in the physical body. With such an action, the "animal spirits," which are congregated in cavities around the brain's pineal gland, would receive the message from the mind, then morph into some sort of ethereal matter and travel through the bloodstreams to cause the body to act in accord with the message's directive. A physical sensation would be the reverse of this process.¹⁸

¹⁸ Descartes' discussion of animal spirits appears numerous in his various works. Two prominent references are in *Discourse on Method and Meditations on First Philosophy*, Part Five (see above reference) and

Descartes' incredible "animal spirit" explanation just didn't satisfy his critics. Eventually efforts to explain dualistic mind/body causal interactionism gave way to a more "scientific" mechanism, a thoroughgoing materialism that ignored if not entirely negated the spiritual soul. But with the loss of the soul, the person is lost as well. The immaterial and immortal soul as the locus of intellect, free will and the very core of one's individual identity, all of which are crucial to what defines us as persons disappeared.

After Descartes' and others' efforts to explain how the mind/soul can causally interact with the body were rejected within the mechanism of modern science, and the nature of the human person was reduced to nothing more than the physical matter which comprised the human being. All of the actions of human intellect, human will, and bodily motility were reduced to a neural network of bioelectrical cause/effect physical impulses. Computer AI, which is also basically a network of electrical impulses, could then be compared with human intelligence, though presently not as complex and sophisticated a network as is characteristic of humans. For many if not most AI innovators, this difference is not, however, a matter of a difference in kind but is merely a matter of a difference in degree: Given time and continued development AI will catch up to and perhaps even surpass human minds.

Fortunately, there is hope that the missing human person within the culture of scientific mechanism and AI development can and will be found and respected. The Vatican has a strong interest in ongoing evaluation and guidance of AI development from the perspective of personalism. One leader in this effort is Rev. Philip Larrey, the Chairman of Humanity 2.0. Fr. Larrey argues that Catholic teaching offers a vital framework for assessing the promise and perils of AI.¹⁹ He warns, for example, that so-

Passions of the Soul, Translated and Annotated by Stephen Voss (Indianapolis: Hackett Pub. Co. 1989) Articles 4, 5, 6.

¹⁹ The following quotes and interpretations, paraphrase and highlight material from an extensive interview with Fr. Philip Larrey by Joan

called “Singularity,” the union of AI with biologically-based intelligence, that developers predict will happen around 2045, is a possibility that must be challenged. Fr. Larrey believes that the fundamental Aristotelean and Thomistic framework of anthropology, cosmology and human nature is excellent for approaching the weighty issues that current AI development is creating. His leadership of Humanity 2.0, which is endorsed by the Vatican aims to construct “guardrails” to prevent bad outcomes with the use of AI. These guardrails are anchored in the framework of the Catholic tradition. Humanity 2.0 thereby aims to establish strong ties with the leading AI innovators and facilitate “collaborative ventures between the traditionally siloed public, private, and faith-based sectors.”

Some of the fundamental principles upon which the guardrails are built and collaborative ventures are enacted are: 1) Human beings are made in the image and likeness of God. 2) Human beings are a substantial unity of soul and body, matter and form. This traditional hylomorphism is of the very composite nature of the human person. 3) The human soul is immortal. Fr. Larrey avers that this principle establishes very important guardrail because of the life extension claims and plans within the AI field. For example, the futurist Ray Kurzweil believes that technology will allow people to rejuvenate themselves enabling them to live forever. But human immortality depends on God’s plan for the human person. Also, radical life extension evokes the question of the meaning and purpose of human existence, an issue for which the Church has a rich tradition of wise doctrine. These three principles serve to identify and preserve the human person and sublimely express the Christian purpose of life, namely, to know, love and be with God in this life and the next.

Fr. Larrey’s work also advocates for what is called today “human-centered AI,” which aims to keep the focus on the

Frawley Desmond entitled “Human-Centered AI: How should the Church Engage with Emerging Technologies?” *National Catholic Register*, July 30 – August 12, 2023, vol. 99, No.17.

human person in the operational uses of AI technologies. For instance, that focus can help us understand that what AI, like Chat GPT-4, can do is not the same as human intelligence. Some of the essential differences between human intelligence and AI are indeed differences in kind such that AI will never, even with ongoing development, catch up to human intelligence.²⁰

1. Tasks that are easier for AI are those that require simple rules calculations. Tasks that are easier for humans are those that require thinking, creativity, and critical reasoning.
2. Human intelligence is the only type of intelligence that can improve itself over time. AI can improve over time only by receiving new data ultimately from human sources.
3. Human intelligence relies on natural cognitive abilities such as intuition, creativity, empathy etc., which AI cannot emulate completely or replicate accurately. AI relies mostly on data processing power and rule-based systems (programming) which are built and defined by human users.
4. Human brains are unique in that they contain billions of interconnected neurons which work together to form neural networks. Humans by the intentionality of their wills can direct the use of these networks so that there results a kind of intelligence synergy which can integrate sensory cognition, emotion, and relevant data. The result of the intelligence synergy is a “whole” which is greater than the sum of its parts. That “whole” cannot in any way be reduced to the mechanistic cause/effect interaction of all of the various parts. AI systems typically consist of processing units or “neurons” which rely on complex algorithms so that processors can be specialized for recognizing objects or predicting outcomes. The causal role and efficacy of each part of such recognition or prediction can be identified so that its mechanistic status can be explained.

²⁰ This discussion of differences in kind between human intelligence and AI paraphrases, quotes from, interprets, and goes beyond the articles by Robert Koch: “Importance and Benefits of Human-Centered AI” and

5. “Human intelligence can perform multiple tasks simultaneously while AI based computers excel at performing one task well. Humans can read and understand text while also completing other tasks like writing an email or memo; computers cannot do these things at the same time effectively. In contrast, AI systems are designed to handle multiple tasks simultaneously by breaking them down into smaller parts that can be processed by separate machines.” Again, with AI, a sort of mechanistic reductionism is required to allocate tasks to separate units with each unit having its specific cause/effect operations.

In closing this essay, it must be emphasized that the BIG question, “Can AI entities gain sentience and become persons?”, is a serious question because of three reasons, which reasons can be critiqued thereby mitigating the seriousness of the question. The critiques of these reasons will offer some hopeful strategies for resisting and possibly, to some extent, undermining the culture of mechanism which is pervasive today. Given how bleak the consequences of AI development without being checked by guardrails are, it is worthwhile to conclude this essay with a legitimate hopeful message.

1. AI continues to be more and more anthropomorphically valenced. The more AI is packaged as a human person the more we are likely to think of it as such. But, if we maintain a conscious awareness that even though the entity is acting like a human being, that is only its packaging. What is beyond that is nothing more than a sophisticated machine. We can not let ourselves be fooled or drawn in by anthropomorphized behavior, appearances, or clever conversational ploys which can influence us to believe that the entity is a type of human being.

“Human Intelligence vs. AI.” The articles can be found respectively at: <https://www.clickworker.com/ai-glossary/human-centered-ai>; and <https://www.clickworker.com/customer-blog/human-intelligence-vs-artificial-intelligence-human-ai>

2. The culture of mechanism, which dominates cyber development and the natural, social, and behavioral sciences in general, aims to erase the human person. It denounces the claim that a human person possesses a spiritual soul as unscientific and affirms that only mechanistic science is real. Mechanistic science denies any sort of immaterial, incorporeal facet of human being. A very significant starting point in resisting mechanistic culture is realizing what the agenda of that culture is. We cannot accept that AI systems and their robotic embodiments are human unless we ignore the human soul. For example, we need to be educated in the differences in kind between AI and human intelligence. The in-kind differences are signs that human intelligence, an application of the faculty of our intellect which is an essential property of our soul, is not entirely reducible to mechanistic material brain activity.
3. Finally, within the culture of mechanism we are not just losing but maybe have even lost an understanding of what the soul is and what it means for us as human persons. Our soul constitutes the core of our substantial identity as composite persons. Our soul, with its spiritual faculties of intellect and free will, is the source of our freedom, rational intentionality, self-governance, our ability to conceptualize, our capacity to know right from wrong, and ultimately our civil liberty and moral accountability. Without an understanding of our soul, we are unable to affirm our free will, our distinct rationality, and our moral agency. The soul today in our mechanistic culture is, sadly, if it is recognized at all, some sort of nebulous spiritual/emotional center of who we are as persons. It might as well be a bundle of emojis wrapped in some ethereal gauze. Relying on Christian tradition, educating ourselves in the teachings of such groups and movements as Humanity 2.0 and Human-Centered AI can certainly provide a hopeful direction for recovering our dignity and for finding our lost souls as human persons. ■

Artificial Intelligence and the Culture of Mechanism

SUMMARY

Artificial Intelligence (AI) programs, systems and robots are currently surrounded by controversy. Even some of the well-known innovators of AI, such as Elon Musk, are advising caution with the headlong rush to develop AI with greater degrees of sophistication. Still, there is no doubt that AI has become practically ubiquitous today and it seems that without, for instance, our Siri and Alexis, we just wouldn't be able to get along as well as we do in our social lives, our work, our personal security, and our home financial transactions. But the problems with AI, the real, the hypothetical and the philosophical are daunting challenges, which unfortunately are not being taken seriously enough. After an extensive critique of many of these AI problems that have occurred and could very well occur soon, this study focuses on the BIG question of AI, namely: Can or will AI become sentient, develop a consciousness? The position advanced is that this question presumes that the traditional understanding of the human soul, the very core of our being as human persons, is actually nonexistent and AI is evolving rampantly within a culture of mechanism that dominates the social, behavioral, and natural sciences. Within such prevailing mechanism there is no doubt that AI will be dubbed as conscious, another "human" species. Only a penetrating philosophical critique of the flaws in AI mechanism which relies on such movements as Human-Centered AI can hope to save the "lost human soul" and find the "missing human person" in today's world.

Keywords: Artificial Intelligence, mechanism, Rene Descartes, technological valence, human-centered AI, Elon Musk, AI dangers, addiction to social media, David Hanson, Fr. Philip Larrey, personalism, AI vs. human intelligence, the lost soul in AI

BIBLIOGRAPHY:

- Associated Press, “Experts Fear Technology’s Ability to Take Election Misinfo to a New Level,” May 15, 2023, *Wheeling Intelligencer*.
- Busch, Corlann Gee, “Women and the Assessment of Technology”, in: *Technology and the Future*, ed. by Albert Teich, St. Martin’s Press, New York 1997, 161–162.
- Bradley, Omar, Quote from Goodreads, https://www.goodreads.com/author/quotes/5764641.Omar_N_Bradley.
- Chatfield, Carly, “AI Gone Rogue: Six Times AI Went Too Far,” December 2021, <https://www.makeuseof.com/ai-gone-rogue-times-ai-went-too-far>.
- Descartes, Rene, *Discourse on Method and Meditations on First Philosophy*, translated by Donald Cress, Hackett Pub. Co. Indianapolis 1980, 29–31. Also, *Passions of the Soul*, translated by Stephen Voss, Hackett Pub. Co., Indianapolis 1989, Articles 4, 5, & 6.
- Desmond, Joan Frawley, “Human-Centered AI: How Should the Church Engage with Emerging Technologies?,” Interview with Fr. Philip Larrey, *National Catholic Register*, July 30 – August 12, 2023, vol. 99, No. 17.
- Donohue, William, “Why Are Young People So Lonely?,” in: *Catalyst: Journal of the Catholic League for Religion and Civil Rights*, vol. 50, No. 5, June 2023.
- Eskovitz, Nir, “AI isn’t close to becoming sentient – the real danger lies in how easily we’re prone to anthropomorphize it,” March 15, 2023, <https://theconversation.com/ai-isnt-close-to-becoming-sentient-the-real-danger-lies-in-how-easily-were-prone-to-anthropomorphize-it-200525#>.
- Harfield, Jake, “Is AI Dangerous? Five Immediate Risks of AI,” June 2021, <https://www.makeuseof.com/is-ai-dangerous-5-immediate-risks-of-artificial-intelligence/>.
- Koch, Robert, “Importance and Benefits of Human-Centered AI,” <https://www.clickworker.com/ai-glossary/human-centered-ai/> and “Human Intelligence vs. AI,” <https://www.clickworker.com/customer-blog/human-intelligence-vs-artificial-intelligence-human-ai/>.

- Lam, Raymond, "The Potential of Personhood: David Hanson on How AI and Human Beings Can Help Each Other." April 27, 2023, <https://www.buddhistdoor.net/features/the-potential-of-personhood-david-hanson-on-how-ai-and-human-beings-can-help-each-other/>.
- McDonald, Matt, "The Artificial Intelligence Educational Revolution," *The National Catholic Register*, September 10–23, 2023.
- Metz, Cade and Schmidt, Gregory, "Elon Musk and Others Call for a Pause on AI," *New York Times*, March 29, 2023, <https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html>.
- Phillips, Joshua, Letter to the Editor, April 29, 2023, *Wheeling News Register*.
- "Social Media Addiction Statistics," June 1, 2023, <https://truelist.co/blog/social-media-addiction-statistics>.
- Teich, Albert, Editor, *Technology and the Future*, Introduction to the Sixth Edition, St. Martin's Press, New York, 1997, iv–v.
- Tegmark, Max, "How to Make AI the Best Thing to Happen to Us," <https://www.npr.org/sections/13.7/2017/10/08/555044340/how-to-make-ai-the-best-thing-to-happen-to-us>.