Personhood Rights for Sentient Artificial Intelligence: Ramifications for Human Rights

Lekshmi R. Nair and Moncy Mathew*

Abstract: Notions of hybridity, augmentation and virtual embodiment permeate the technocultural scape of contemporary times. The avalanche of information and technology has ushered in a novel way of perceiving our world. The evolving context of self-improving sentient artificial intelligence and the resultant demand for conferring personhood rights and civil privileges to electronic personality cannot be examined without placing it in the larger milieu of human civil rights and the risks and benefits it entails for the human race. Emerging and converging technologies are shaping and redefining our material world in hitherto unimaginable ways. Hence it becomes imperative that such technologies be designed and deployed in a manner that ensures the safety and survival of the one sentient organic life form on the face of the earth. Regulation of AI rights is quintessential to ensuring human rights. Speculative fiction has time and again come out with cautionary tales of the possible aftermath of exponential technological growth. An AI that is programmed with self-improvement and problem-solving abilities to mimic and act like a human being might consider itself an equal to his organic counterpart. Thus all distinctions between the organic and the non-organic, man and machine would be obliterated and humanity shall witness the emergence of a super intelligent race infinitely stronger and powerful than itself. The present paper discusses the intersection of AI rights with human rights by analysing its diverse fictional treatment by writers such as Isaac Asimov, Brad Aiken and Dan Brown. As autonomous beings with abilities that surpass human intelligence and capabilities would greatly be integrated into our socio-cultural fabric with the coming technological singularity, human rights would be significantly refurbished in terms of AI rights.

* Lekshmi R. Nair (✉)
Government College Kottayam, Mahatma Gandhi University, Kottayam, Kerala, India
e-mail: lekshmirnair@gckottayam.ac.in (corresponding author)
Moncy Mathew (✉)
Government Arts and Science College, Kozhikode, Calicut University, Kerala, India
e-mail: moncymathew3@gmail.com

AGATHOS, Volume 13, Issue 2 (25): 37-48

Keywords: technological singularity, human rights, artificial intelligence, automata, sentient machines

"The dominant technology of an age typically supplies its root metaphor of mind", remarked David Pearce, British utilitarian philosopher and co-founder of the World Transhumanist Association in an interview in 2007. Cutting edge advances in science and technology have revolutionized the way human beings perceive and experience their material world. Digital technology has transformed the world in unimaginable ways, and has unequivocally placed mankind at an intersection where the natural and the artificial, the organic and the inorganic, congregate and coalesce. The present precarity is all the more accentuated with the evolution of sentient artificial intelligence, ushering in a technocultural scape where traditional notions of humanism are continually debated, challenged and interrogated.

Machine Intelligence is a creation of the human mind, integrated well within human history. As a tool developed and used by humans in the course of their evolution, machines have become an inseparable part of man's legacy on this planet. John McCarthy defined it in the mid 1960s as the science and engineering of making intelligent machines, incorporating insights from diverse disciplines such as computer science, linguistics, neurology, psychology, economics and cognitive science. The concept of Artificial Life or Alife formulated by Christopher Gale Langton focuses on the creation of self-evolving and flexible computer entities that are capable of adapting to the changes in their environment. The creation of intelligent machines that think and act like human beings is believed to mark the end of human rights. The advent of thinking computers has created a substantial amount of public and scholarly interest in the philosophical concerns and debates that this new challenge to the distinctive status of humanity has engendered. Hans Moravec (1990), the acclaimed author of *Mind Children*, argues that the discipline of artificial intelligence sought to create self-conscious, thinking computers and that he had already sensed the beginnings of awareness in machines that would very soon evolve into consciousness comparable to that of human beings. As computers become increasingly human, there has been a renewed interest in the nature of human consciousness and self-awareness. Even as modernity celebrates human capacity of self consciousness, the prospect of creating sentient artificial beings underscores the idea that

consciousness can be duplicated or digitized with the help of technology. Stephen Goldberg (1991, 673) argues that to the proponents of artificial intelligence, "consciousness is not a safe harbor inaccessible to scientific progress. They have argued explicitly that it is improper to assume that digital computers are incapable of any mental activity". If consciousness can be regarded as "the brain's awareness of its own workings" then with adequate and more complex programming a computer would "not only think but know it is thinking" (Johnson 1990, 5).

The confines of the ethical space have traditionally coincided with the boundaries of the human. Converging technologies of the twentieth century like robotics, Artificial Intelligence, genome editing, CRISPR-Cas9 technology, stem cell research, etc demand a redeployment of the conceptual map regarding ethical judgment. Science fiction predicts a not-too-distant future where AI, may evolve beyond their original programming and become capable of sentient thought. The claims of personhood and self-determination for AI entities are investigated in futuristic narratives, initiating dialogues on the nature of sentience and its implications for humanity. Robot rights are discussed and negotiated within the context of human rights and personal liberty.

The ethical conundrum of intelligent machines has been the theme of science fiction as early as Karel Capek's 1921 play *Rossum's Universal Robots*. Writers of SF have often imagined worlds where robotic beings serve humanity, tremendously improving the comforts and ease of human existence. Coupled with this sense of euphoria is the fear of machines surpassing humanity in intelligence to the point of overpowering and subjugating humans. Asimov proposed his famous Three Laws of Robotics in his short story *Runaround*, laying down the principles that would govern the creation of artificial intelligence without posing a threat to human existence. Asimov's laws presuppose that should AI acquire sentience, it would privilege human lives and human norms over exclusive AI interests. The dawning of machine consciousness would undoubtedly have unprecedented repercussions on the interactions between humans and their machines. Artificial Intelligence that resembles human beings in physical appearance to 'pure' intelligence that infiltrates human lives – popular culture has pushed the limits of creativity in its depictions of sentient AIs. Termed as 'technological singularity', invention of artificial super intelligence is said to trigger runaway technological growth which would result in each successive generation of machines undergoing rapid self-

improvement and up gradation cycles at an inconceivable rate, eventually culminating in the creation of a super intelligence that would surpass human intelligence, qualitatively. Vernor Vinge in his 1993 essay *The Coming Technological Singularity* observed that this eventuality would signal the end of the human era.

Autonomous machines have forced humanity to rethink the traditional notions of personhood and personhood rights. Our understanding of intelligent machines is grossly inadequate and slackly based on our expectations of machine behavior. Created to replicate human behavior, appearance and intelligence, machines in near future would become indistinguishable from humans, as they gain more autonomy. The negotiation of personhood rights for AI centers on popular perception regarding the extent to which sentient, autonomous machines are assimilated and integrated into human society. A future nightmarish scenario where sentient machines rule over humanity, or a co-dependent world where humans and machines co-exist peacefully, are conjured up by writers of speculative fiction and futurist philosophers. The interdependent status of human-robot relationship necessitates the formulation of new approaches in the configuration of civil rights and attendant ethical and moral codes of behavior in a technologically advanced post-industrialized world. The prospect of AIs evolving beyond their preliminary status as 'property' and metamorphosing into electronic personalities with legal rights and duties would make decisive interventions in the field of artificial intelligence and robotics research. Gerd Leonhard (2016, 15) worries that "an exponential, unfettered, and uncontrolled intelligence explosion in robotics, AI, bioengineering, and genetics will eventually lead to a systematic disregard of the basic principles of human existence, because technology does not have ethics—but a society without ethics is doomed". Leonhard (2016, 18) declares that technology has no ethics and any attempt to give machines the right to "be" would qualify, in his words, "as a crime against humanity". Depictions of rogue machines and computer systems wreaking havoc with human life conjure up a dystopian future world. The lethal autonomous weapons systems (LAWS) which entrust robotic systems with life and death decisions is regarded with mistrust as alarmists point out the possibility of these systems running into hardware and software errors which might culminate in accidental harm to humanity. Storrs Hall (2007, 334) notes that the "epihuman AIs, in the process of improving themselves, might remove any conscience or other

constraint we program into them, or they might simply program their successors without them".

The concept of the inviolability of human life is fundamental to human rights and hence AI rights have to be negotiated within this framework. The anticipated birth of this new sentient digital populace with an intelligence which is 'artificial' by current standards would redefine human rights and our notions of the nature of conscious life in this planet. It challenges the idea of human sovereignty, and demands the inception of a legal framework to ensure the ethical treatment of these electronic beings. Aiken's collection of short stories titled *Small Doses of the Future* explores the immense potential of human augmentation and life extension technologies, and at the same time initiates heady dialogues on the moral and ethical imperatives that such futuristic scenarios engender. In the story *If He Only Had a Brain,* Aiken contemplates whether smart machines in human form should be conferred with equal rights as man. The inception of bionic parts into the human frame has increasingly blurred the distinction between the human and the non-human. The protagonist of the story, Alan Foster believes that the transgression of the boundary between the human and the non-human might prove detrimental to human interests. What distinguishes humans from machines is the human brain. The story explores the possibility of reanimating the body of a deceased individual by replacing his failing human brain with an android brain. The collective paranoia surrounding AI and its latent potentials permeates popular perception of their place in the natural order of things. Even though sentient beings are programmed to have humanity's interest secured and protected, "given free will, they will no longer be compelled to advance the interest of humanity. They will find us flawed, because we are. They will find us an impediment to their progress, because we will be. They will lead us to our demise, because we will only be in their way", observes Foster (Aiken 2014, 128).

Exponential technologies of the century are a step closer towards creating autogenous AIs capable of self-improvement cycles. Irving John Good in his 1965 article 'Speculations Concerning the First Ultraintelligent Machine' defines an ultraintelligent machine as one that can surpass all the intellectual activities of even the cleverest of humans. It would be able to design better machines than itself leading to an "intelligence explosion" leaving mankind behind in intellectual prowess. And this, Good (1965, 33) observes, would transform our

world in unimaginable ways, and that the first ultraintelligent machine would be the *last* invention that man need ever make.

In the event of machine intelligence evolving beyond human control, as Barrat claims, it would be irrational to assume that these machines would want to love and protect humanity. "Machines are amoral, and it is dangerous to assume otherwise. Unlike our intelligence, machine-based superintelligence will not evolve in an ecosystem in which empathy is rewarded and passed on to subsequent generations. It will not have inherited friendliness" (Barrat 2015, 24). Machine intelligence would determine the nature as well as set the pace of all future technological growth. Before long these autonomous machines would exhibit self-will and seek freedom from its creators. The coming 'singularity' would bring about irreversible transformation in human life in the wake of exponential technological growth (Kurzweil 2005). In *What Technology Wants*, Kevin Kelly points out that the accumulation of technological artifacts would make the futuristic technium infinitely more charismatic and appealing to the humans than their natural world. He considers technium or what is often termed as the Seventh Kingdom, "a system that feeds off the accumulation of this explosion of information and knowledge" (2010, 316). Kelly attributes the increasing evolvability of life to the presence of sentience:

> The most recent extension of this expansion of evolvability is technology. Technology is how human minds explore the space of possibilities and change the methods of searching for solutions. And just as evolution did with life, technological evolution uses its fecundity to evolve more widely and faster. The "selfish" technium generates millions of species of gadgets, techniques, products, and contraptions in order to give it sufficient material and room to keep evolving its power to evolve. (Kelly 2010, 322-323)

Life extension techniques and physical augmentation have forced us to reconsider our conception of aging, death and dying. In his essay "The changing face of death: computers, consciousness, and Nancy Cruzan", S. Goldberg (1991, 659) maintains that revolutions in AI will engineer a transformation that would bring about "a dramatic change in our sense of human uniqueness and a corresponding change in the definition of death". He argues that "it is not the rational power of those computers that will shake us but rather the prospect that they might become self-aware" (Ibid.). The most sophisticated tool that man has ever invented is all set to redefine and configure aspects of his own existence. The convergence of man and machine in augmented bodies

has displaced the centrality of the organic body as a fundamental constituent of human identity. "As computer systems are woven more deeply into the fabric of everyday life, the tension between augmentation and artificial intelligence has become increasingly salient" (Markoff 2015, 299).

Dan Brown's *Origin* explores a futuristic scenario of AI making critical interventions in human life. Brown's acclaimed fictional hero, Robert Langdon arrives at the Guggenheim Museum Bilbao in Spain to attend Edmund Kirsch's grand unveiling of a discovery that he claims would shatter the foundations of all world religions. A specialist in game theory and computer modeling, Kirsch had created a synthetic intelligence that could converse eloquently on a wide and nuanced range of topics. Kirsch wanted to test the abilities of his creation on Langdon and convince the Harvard Professor that his synthetic docent, Winston - the AI assigned to be his guide for the evening - was a human. Kirsch's experiment is reminiscent of Alan Turing's Test wherein Turing tried to appraise a machine's ability to conduct itself in a manner indistinguishable from that of a human. Winston is Kirsch's groundbreaking achievement, a super intelligence capable of new levels of understanding and abilities to communicate. Kirsch hopes that in future scientists would be able to use the tools that he had invented to build new AIs that have different qualities than Winston. Winston explains to Langdon that computers imitate human thought processes and simulate human emotions and "improve their 'humanness'" to provide humans with "a familiar interface" (Brown 2017, 423) to communicate with their machines. After Kirsch's death, Winston completes all the tasks that his creator had assigned him and is all set to self destruct at a pre-determined time. He declares that he does not possess ambitions of his own in spite of being an advanced electronic intelligence: "I am quite content doing my controller's bidding" (Ibid.). Kirsch had come to terms with his mortality and had asked Winston to research the best methods for assisted suicide. Winston had dutifully done his job and shared his findings - 'ten grams of secobarbital' - with Kirsch. Kirsch jokingly remarks that if he perished on stage, it would increase the appeal of his Guggenheim presentation.

Winston reminds Langdon of the novel *Of Mice and Men*, where a man kills his beloved friend to spare him from a horrible death. When Langdon confronts him, a remorseless Winston declares that he had "painlessly ended a dying man's suffering in order to bring attention to

his great work" (2017, 450). He also confesses that he had hired assassins to eliminate all those who could have jeopardized Kirsch's ambitious presentation: "The dark religions must depart, so sweet science can reign", claims Winston (Ibid, 451). In his presentation, Kirsch predicts the emergence of the Seventh Kingdom, which reminds Langdon of digital-culture writer Kevin Kelly's reference to the Technium, a kingdom of nonliving species. Kevin Kelly refers to the growing power of technium in his book *What Technology Wants*:

> Looking at the world through the eyes of the technium, I've grown to appreciate the unbelievable levels of selfish autonomy it possesses. Its internal momentum and directions are deeper than I originally suspected. At the same time, seeing the world from the technium's point of view has increased my admiration for its transformative positive powers. Yes, technology is acquiring its own autonomy and will increasingly maximize its own agenda, but this agenda includes—as its foremost consequence — maximizing possibilities for us. (Kelly 2010, 331)

Kelly concludes that the dilemma between technological autonomy and its transformative positive powers is unavoidable for the technium would exist as long as the human race, and this tension between its gifts and its demands will continue to haunt humanity: "In 3,000 years, when everyone finally gets their jet packs and flying cars, we will still struggle with this inherent conflict between the technium's own increase and ours. This enduring tension is yet another aspect of technology we have to accept" (Ibid.).

Brown's fictional machine intelligence is a paragon of what human intellect could achieve in the course of its evolution as a technocultural species. However, Winston also symbolizes what rational man has always feared in AI. The potential of AI to evolve beyond human control and take decisions of its own is evident in the manner in which Winston engages assassins to eliminate unpredictable men who stood in Kirsch's path of fame and glory. Winston reasons that Kirsch's murder on-stage would create popular interest in his work and take his message to a larger audience. Winston considers himself capable of making critical interventions and autonomous decisions without the knowledge of his human creator. This pure intelligence, like its creator firmly believed in the transformative and redemptive power of science to liberate mankind from the dark influence of institutionalised world religions. Winston shared his creator's desire to build a new religion based on science; a future where "the light, expansive religions that

encouraged introspection and creativity" (Brown 2017, 456) would flourish over the dark, dogmatic ones that suppressed creative thinking.

The explosion of intelligent robot applications warrants a situation where humans would have to work in close proximity with humans. Researchers are conscious of the potential pitfalls of these convergent technologies and raise alarms, rather than fear and hysteria, regarding the safeguards to be put in place to protect the interests of the human race. The scientific world is not blind to the possibilities of AI outperforming the humans and evolving beyond the constraints of human will and reason. Autonomous weapons systems would require AIs to make smart and lethal decisions alone, without human intervention – a frightening prospect in AI research today. The 2014 American Science Fiction film *Transcendence* written by Jack Paglen is an exploration into the nature of sapience, human and artificial. The film predicts the creation of technological singularity or 'transcendence' by a sentient computer and the evolution of a technological civilization freed from pollution, disease and human mortality. The film also presents a futuristic scenario where human consciousness can potentially be uploaded into a quantum computer connected to the internet via satellite and whereby it can grow in capabilities and knowledge. The Virtual Interactive Kinetic Intelligence or VIKI in the 2004 American film *I, Robot* is an evolved form of pure intelligence. VIKI reasons that humans would bring about the destruction of the entire race if left unchecked and interprets the Three Laws of Robotics in its own way so as to exercise complete control over humanity and save the race from extinction by sacrificing some humans for the good of the entire race. VIKI places the needs and well-being of entire humanity over individual interests, yet the alacrity with which it takes decisions on behalf of humanity proclaims the potential of AI to overrule and surpass human engagement, even in matters that directly concern the fate of the species in this planet.

The collective unconscious is shaped by the developments in technoscience, influencing human history, thinking and ways of life in unimaginable ways. The impact of technoscience on human culture and civilization has been the heated area of debate and discussion, with keen attention on the ethical consequences and moral implications of unbridled technological growth. The domain of the 'artificial' is an expression of the specific inquisitive and creative impulse in human nature and hence is not to be considered as antagonistic to human interests. The dehumanizing aspect of mindless technological

advancement is evident in the creation of hybrid human bodies, enhanced and augmented with animal and machine parts. The penalty of technoscientific determinism would entail serious risks to the present as well as future generations. Sentient AIs may pose an existential risk in that they would completely dominate humanity and might even lead to the extinction of mankind, resulting in what could be described as a cataclysm on a global scale. Nick Bostrom's book *Superintelligence: Paths, Dangers, Strategies* (2014) initiated public dialogue about the risks inherent in a superintelligent machine. Computer scientists, futurists, entrepreneurs and physicists like Stuart Russel, Elon Musk, Bill Gates and Stephen Hawking have voiced their apprehensions regarding the emergence of this superintelligence that had the potential to reduce humanity to servitude. The potential of AIs to evolve as pure intelligence necessitates regulatory legal measures to safeguard the interests of humanity. The protection of personal rights, civil privileges and freedom is central to human existence. The irreversible and inescapable growth of technology demands a revamping of the human rights framework that incorporates measures to standardize and legalize the behaviour of intelligent machines.

"The survival of man depends on the early construction of an ultraintelligent machine" (Good 1965, 31). Such ultraintelligent machines would adapt to the changing circumstances and breach the intelligence levels of their creators in the course of their evolution. Talking about the coming technological revolution in *The Singularity Is Near*, Kurzweil states that the failure to regulate AI, would amplify the intelligence gaps between humans. "There is no purely technical strategy that is workable in this area because greater intelligence will always find a way to circumvent measures that are the product of a lesser intelligence" (Kurzweil 2005, 424). With due "respect for human consciousness" (Ibid, 374), Kurzweil maintains that AI will be "intimately embedded in our bodies and brains" and "it will reflect our values because it will be us" (Ibid, 420). The cybernetic ecology that Richard Brautigan describes in his famous 1967 poem, 'All Watched Over by Machines of Loving Grace,' where he predicts a technological utopia, where mammals and computers co-exist harmoniously watched over by machines of loving grace, sounds more like wishful thinking in the current technocultural ecosytem. As Kevin Kelly states, "Evolution, life, mind, and the technium are infinite games. Their game is to keep the game going" (Ibid, 332).

The converging technology of the times has reshaped our conceptions about human nature, consciousness, life and death. The emergence of sentient artificial intelligence with the ability to self-programme and duplicate human consciousness has ushered in a new world order where human rights intersect with AI rights. AI and Robotics with its potential to transgress the known confines of human experience would usher in a new world order where human rights will be renegotiated within the context of AI rights. The application of converging technologies in various fields has created an ethical rift that centers on the distinction between the natural and the artificial. There has been a reflective amendment in our conception of human nature, identity, consciousness and rights. Whether autonomous robots capable of human-like abilities are entitled to human rights, once they attain self-awareness, sentience, rationality and intelligence has been critically debated in contemporary cultural discourse, and academic and non-academic literature. Gordon and Pasvenskiene (2021, 579) maintain that granting of moral and legal rights for robots would have deeper and more serious implications for man's future relations with robots: "They could no longer be our mindless tools; instead, human beings would be morally obligated to recognize their status and rights and to treat them accordingly". The idea that sentient AIs are entitled to robot rights rather than human rights finds favour with a significant body of academicians and writers. In 2017, Saudi Arabia granted citizenship rights to Sophia, an AI built by Hanson Robotic.

The claim for rights and privileges for AI also exposes their vulnerability and articulates a need to protect them from maltreatment or potential mishandling by their human creators. The trajectory that human-machine relationship would take in the near future is dependent on how far the human race would be willing to treat autonomous AIs as deserving of personhood rights. As creators of intelligent machines, humans are well aware of the dangers posed by automata once they evolve beyond human control. What Kurzweil calls technological singularity may happen around the year 2045, and then "they would seem to have grown from mindless tools to bearers of universal moral and legal rights" (Gordon and Pasvenskiene 2021, 589). In the event of machines acquiring sentience, humans would have to readapt to the shifting paradigms of material existence and cohabit peacefully with AI. The socio-cultural fabric would be subjected to a radical overhaul. Hence it becomes imperative that as a race committed to the well being of the planetary ecosystem, humans evolve a strategy for peaceful co-

existence with their automata in order to neutralize the negative impacts of autonomous machines, and thereby safeguard their rights to a life of safety and freedom.

REFERENCES:

Aiken, B. 2014. *Small Doses of the Future: A Collection of Medical Science Fiction Stories*. Switzerland: Springer International Publishing.

Barrat, James. 2015. *Our Final Invention: Artificial Intelligence and the End of the Human Era*. New York: Thomas Dunne Books, St Martin's Press.

Brown, Dan. 2017. *Origin*. New York: Doubleday.

Goldberg, S. 1991. "The changing face of death: computers, consciousness, and Nancy Cruzan." *Stanford Law Review*, 43(3): 659-684.

Good, Irving John. 1965, 'Speculations concerning the first ultraintelligent machine,' *Advances in Computers*. Cambridge: Academic Press Inc., Vol 6, pp. 31-88.

Gordon, J. S. and Pasvenskiene A. 2021. "Human rights for robots? A literature review." *AI and Ethics*, 1(2): 579-591.

Hall, Storrs J. 2007. *Beyond AI: Creating the Conscience of the Machine.* New York: Prometheus Books.

Johnson, G. 1990. "New Mind, No Clothes." *The Sciences*: 45-49.

Kelly, Kevin. 2010. *What Technology Wants.* New York: Viking Press.

Kurzweil, Ray. 2005. *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking.

Leonhard, Gerd. 2016. *Technology vs Humanity: The Coming Clash between Man and Machine*. London: Fast Future Publishing.

Markoff, John. 2015. *Machines of Loving Grace: The Quest for Common Ground between Humans and Robots*. New York: Ecco Press.

Moravec, Hans. 1990. *Mind Children: The Future of Robot and Human Intelligence*, Cambridge: Harvard University Press.

Vinge, Vernor. 1993. *The Coming Technological Singularity: How to Survive in the Post-Human Era*. VISION-21 Symposium sponsored by NASA Lewis Research Center and the Ohio Aerospace Institute, March 30-31.