AI and Ethics
https://doi.org/10.1007/s43681-022-00170-8

ORIGINAL RESEARCH

Normative ethics, human rights, and artificial intelligence

Shailendra Kumar[1] · **Sanghamitra Choudhury**[2,3,4,5,6]

Received: 28 March 2022 / Accepted: 27 April 2022

Abstract
At some point in the future, nearly all jobs currently performed by humans will be performed by autonomous machines using artificial intelligence (AI). There is little doubt that it will increase precision, comfort, and save time, but this coincides with the introduction of many ethical, social, and legal difficulties as well. Because machines will be performing all of the tasks that humans used to, they cannot be kept exempt from the ethical principles that humans follow. However, because digital machines can only understand 0 and 1, encoding complex philosophical ideas in 0 and 1 would be an assiduous task. These great difficulties offer the opportunity to revisit some of the basic and time-tested normative moral theories advanced by modern philosophers. There could be significant advantages for the many players in AI, namely producers and consumers, thanks to these moral philosophies. Customers could use it to make a purchase decision concerning AI machines, whereas manufacturers could use it to write good ethical algorithms for their AI machines. To handle any ethical difficulties that may develop due to the use of these machines**,** the manuscript will summarise the important and pertinent normative theories and arrive at a set of principles for writing algorithms for the manufacture and marketing of artificially intelligent machines. These normative theories are simple to understand and use, and they do not require a deep understanding of difficult philosophical or religious notions. These viewpoints claim that good and wrong may be determined merely by applying reasoning and that arriving at any logical conclusion does not necessitate a thorough understanding of philosophy or religion. Another goal of the manuscript is to investigate whether artificial intelligence can be trusted to enforce human rights and whether it is right to code all AI machines with one uniform moral code, particularly in a scenario where they will be doing different jobs for different parties. Is it possible to use the diversity of moral principles as a marketing strategy, and could humans be allowed to choose the moral codes for their machines?

Keywords Artificial intelligence · Ethics · Human rights · Moral philosophies · Normative ethics · Categorical imperative

1 Introduction

The way AI is conceived presents astringent ethical difficulties that are gradually becoming more acute as AI peregrinates from the lab to the market [1]. Expertise is no longer a rare human capital and monopoly, and one can visually perceive AI machines availing and even replacing humans in many highly specialised fields. At present, we can easily see the presence of robots in every field, who not only match but at times even exceed the performance of human beings. In the days to come, these machines are liable to become more and more intelligent and may relieve humans of many conventional chores and errands. In the future, AI machines will be plenarily capable of performing all occupations that currently require highly educated and trained humans, such as doctors, engineers, lawyers, teachers, drivers, accountants, nurses, and even security-cognate occupations such

✉ Sanghamitra Choudhury
schoudhury.oxon@gmail.com

Shailendra Kumar
theskumar7@gmail.com

1 Sikkim Central University, Gangtok, Sikkim, India

2 Bodoland State University, Deborgaon, Assam, India

3 University of Oxford, Oxford, UK

4 Queen's University, Belfast, UK

5 The Hague Academy of International Law, Hague, The Netherlands

6 Jawaharlal Nehru University, New Delhi, India

Published online: 24 May 2022

as police, the army, and detectives. The introduction of the Internet of Things (IoT) has made AI machines a part of the day-to-day lives of human beings. In the IoT, "things" can be anything from humans with implanted heart monitors to animals with biochip transponders to cars with built-in sensors that let drivers know when their tyre pressure is low [2]. Besides benefitting people by giving them total control over their lives, the Internet of Things has also made machines smarter, more autonomous, and more intelligent [3]. As machine learning (ML) becomes more popular, it will eventually be able to aid humans in practically all areas [4]. However, machines will probably make mistakes when performing these specialised and extremely intricate activities. On the one hand, these developments in AI may give humans more time to save their energy and put their focus on other productive thinking. On the other hand, they will also pose some serious challenges, of which one is the fine-tuning of the accountability of AI robots in case anything goes wrong. Tracing responsibility among them would cause substantial ethical and legal concerns. Even if the number of errors made by machines is minuscule in contrast to those made by human specialists, they will indubitably occur. Who would be ethically liable for those errors in those circumstances? Will the company that makes the machines, or the person who uses them, or the person who wrote the machine's algorithm, or the company that sells them, or the government that approved those machines for use, be responsible for the damage caused by those machines? We have seen that, in the case of some lifesaving drugs and vaccines, the government provides immunity to the manufacturers from any damage reported by the drugs or vaccines. An ethical dilemma will naturally occur when these AI machines are forced to pick between two evils. In such a case, which option would these machines select, how would they select it, and why would they select it? How can we use the theories proposed by great moral philosophers to make those AI robots more advanced, practical, ethical, and market-ready? How will these machines address the issue of human rights? Can AI be trusted with human rights? These questions will be addressed in this manuscript. The paper draws on the normative ideas of prominent philosophers for answers to the ethical dilemma of AI robots. These normative theories are straightforward, and they do not necessitate a thorough understanding of complex philosophical or religious concepts to apply them. These views hold that right and wrong may be determined simply by applying logic, and that one does not require a comprehensive understanding of philosophy or religion to arrive at any logical conclusion.

In what follows, the manuscript will cover the history, growth, and development of AI as well as the relevant laws and regulations for robotics, before diving into the ethical, social, legal, and human rights problems raised by artificially intelligent machines. The manuscript will attempt to analyse these challenges through the lens of moral philosophies given by renowned philosophers. In the final section, the manuscript will analyse the challenges in the human rights sector thrown open by AI and follow up with suggestions on how to tackle these challenges.

1.1 History, development, growth, and future of artificial intelligence

The romance of man with machines is not a new-age phenomenon, and one of humanity's most treasured fantasies has been to overcome human imperfections with the help of machines. The ancient Indian scripture, the Rig Veda, which is also regarded as mankind's oldest literature [5], mentions the word *Vimāna*, which is identical to modern-day flying airplanes. Similarly, Daedalus is mentioned in Greek mythology as a brilliant inventor who is credited with creating bizarre gadgets such as man-flying wings and live bronze statues that could cry, laugh, talk, walk, and even perspire. Hephaestus, the Greek God of creation, is credited with creating an animated bronze warrior named *Talos*, who was programmed to guard the island of Crete and sounds eerily similar to the modern-day Robo-Cop of Hollywood films [6]. In 1952, G.R. Josyer claimed to have discovered a Sanskrit text named the *Vaimānika Shāstra*, which goes into great detail about the creation of *vimānas*, or airplanes. The text was authored by *Pandit Subbaraya Shastry* between 1918 and 1923 and is divided into eight chapters and contains 3000 shlokas. When Charles Babbage invented mechanical machines that displayed intelligent behaviour in the late nineteenth century, the platonic love of humans for machines grew more intense and transformed into a torrid affair. Ada Lovelace, often recognised as the first computer programmer, collaborated with Charles Babbage to use algorithms to implement punched cards on his computers. *In 1945, the Electronic Numerical Integrator and Computer, or ENIAC, was introduced in Philadelphia, ushering in the modern computer age. ENIAC, transformed the world's computations from mechanical to digital.* In 1950, Claude Shannon proposed playing chess on computers, which was revolutionary because the game of chess requires a separate intelligence, and computers playing chess demonstrate an advanced upliftment in machine intelligence. Although an American mathematician, James McCarthy, coined the term "artificial intelligence" in 1956, it was the ideal British scientist, Alan Mathison Turing, who laid the conceptual framework for AI in the late 1950s and early 1960s [7]. In 1950, Turing dubbed the "Father of AI," created the famous Turing Test, often known as the "imitation game." In this game, the interrogator is separated from the person or machine being interrogated, and the interaction can only be conducted by a teletype. The interrogator uses a teletype to ask questions, and if the interrogator is unsure whether he

or she is communicating with a person or a machine, the machine is said to think. However, Turing used the phrase "machine intelligence" throughout his life, and the field of AI was not founded until after his death in 1956 at the Dartmouth conference. John McCarthy created a functional programming language designed for AI in 1957. He also invented LISP, or list processing language, which enables one to write flexible programmes that represent basic functions using a list structure. Then there was the "dark age of AI," from 1965 to 1970, when there was almost no effort put into AI, but western fiction writers used their novels to carry forward their AI ideas. However, since the 1970s, when AI scientists attempted to introduce AI through other disciplines such as psychology and philosophy, these difficult days have gained significant traction [8]. J.C.R. Licklider from MIT popularised the concept of an "Intergalactic Network" of computers. That led to the development of the ARPANET (Advanced Research Projects Agency Network), which prepared the groundwork for the modern Internet. ARPANET made TCP/IP the de facto standard on January 1, 1983, and in the years that followed, researchers built the current Internet by cobbling together separate networks. The phrase "Internet of Things" (IoT) was coined in 1999 by British researcher Kevin Ashton, and he described the overall network of interconnected and communicating things linked together like computers do today on the Internet [9–11]. With this type of connectivity, objects can be controlled from a distance. According to Margaret Boden, a prominent figure in the field of AI, intelligence concepts similar to those of AI have long been a philosopher's dream [12]. When quantum computers were invented in 1998, the advancement of computing took a big leap forward. Quantum computers are data storage and processing machines that make use of quantum physics features. The first quantum computer was built by Isaac Chuang of the Los Alamos National Laboratory in 1998. Bits, a stream of electrical or optical pulses representing 1 s or 0 s, are used in today's computers. Qubits, which are subatomic particles like electrons or photons, are used in quantum computers. Heat, electromagnetic forces, and collisions with air molecules can all cause a qubit's quantum characteristics to be lost for the time being. At the same time, qubits can represent a wide range of conceivable 1 and 0 combinations. Quantum computing discoveries are predicted to enhance artificial intelligence as they allow us to expand on what artificial intelligence can already achieve. However, the development of quantum computers is still at a very nascent stage and may take at least a decade to show some promising results [13]. Quantum computing will pave the way for quantum AI in the future, as quantum computers can provide an exponential increase in performance and capabilities for a variety of applications, including AI. The physical system symbol hypothesis of Allen Newell and Herbert A. Simon considers that a physical symbol system has the necessary and sufficient means of general intelligent action [14]. The Dartmouth proposal, adopted in the 1956 summer workshop, is widely considered to be the founding event of artificial intelligence as a field, resolving that every facet of learning, or any other component of intelligence, may be described in such detail that a machine can replicate it [15]. The strong AI hypothesis of John Searle argues that, in the same way, that humans have minds, a properly programmed computer with the necessary inputs and outputs would have a mind [16]. The term "narrow AI" refers to the use of artificial intelligence for a specific job, such as picture recognition, language translation, and autonomous vehicles. At the moment, machines are more precise than humans at these tasks. Researchers want to achieve "artificial general intelligence" in the future (AGI). This would entail intelligent systems that can perform a variety of cognitive activities. These skills, however, are not expected to be realised for decades, according to academics. When it comes to the building of highly astute man-like machines, there have always been two perspectives. One viewpoint encourages us to build a model of the human mind that can be used for this, but the opposing viewpoint says that doing so will make us look like the Darwinian scientist who built Frankenstein when he tried to build a human [17]. This conception of man flirting and romancing with machines and fantasising about their becoming a fundamental part of life, is also well represented in the popular fiction novels of the twentieth century. "*I am solitary and wretched, a man will not consort with me, but one as disfigured and hideous as myself would not gainsay herself to me. My buddy has to be of the same species as me and suffer from the same imperfections. You must construct this thing.*"-Scientist Victor Frankenstein states in Mary Shelley's classic novel Frankenstein [18]. According to Geraci, the creation of intelligent life is simultaneously religious, scientific, and artistic [19]. Homo sapiens is an advanced race because of its superior thinking capacity, and AI has the potential to take it a step further by sanctioning humans to not only understand but additionally develop as sentient beings. But this evolution of human beings from *Dryopithecus* to *Homo Sapiens* should continue and not be confined to the protein-based structure alone. In the next stage of the human evolutionary process, AI should play a greater role. The success of this new human–machine race that will have evolved as a result of the fusion of Homo sapiens and artificially intelligent machines will be crucial to the survival of the human race. Authors would prefer to call it *Artilians*, which is derived from two words: *artificial*, meaning "not natural", and *Ian*, which is a Hebrew baby name meaning "gift from God". On the other hand, scientists such as Stephen Hawking, a giant in the realm of physics whose ideas and opinions were widely respected and taken seriously by the scientific community [20], have expressed

fear that creating highlyintelligent machines that could replicate themselves could lead to the destruction of humanity. According to Hawking, "The emergence of artificial intelligence could herald the annihilation of the human race. It would take off on its own and continue to remodel itself at a rapid pace. Humans, whose biological evolution is slowed, would be unable to compete and would be surpassed" [21].

1.2 AI and normative ethical theories

The apprehension regarding the impact of AI on the humanities and social sciences settled down when Marvin Minsky, one of AI's founding fathers, said, "Only a humanities professor could be so blind to the third exciting possibility: psychology (humanities) could turn out to be like engineering" [8]. This viewpoint by Minsky was outstanding. From the application of AI in the humanities and social sciences, the course of debate has now shifted to writing algorithms that make machines capable enough to deal with ethical dilemmas. AI machines function on a predefined set of algorithms, and while these algorithms may aid in smarting up humans in chess and data processing, developing human-like nous and intuition based on established algorithms will be challenging. Of course, as AI's goals get more sophisticated, learning plays an increasingly important role. Machines equipped with high sensors may assess the situation they are in and logically identify more effective and efficient ways to complete it. Similarly, robots may be trained and coded with normative theories and might figure out how to mix several ideas and use the most appropriate ones. Artificial intelligence researchers might use the data from normative theories to teach machines how to understand and predict everything we want or do not want. For this, the normative theories could either be translated as a set of universal rules that are always valid, or they could also be translated into a set of rules (and, as a result, an algorithm) that can be applied exclusively to specific instances. It's also possible to make a list of scenarios that cover the most common situations, which will cut down on the number of algorithms required. The complexity and sophistication of AI algorithms may also enhance the risk. As a result, modern businesses and technical advancements must be accompanied by proper regulations that regulate the risks connected with them, allowing the AI industry to thrive. At the same time, legislation must allow for enough flexibility so that technological advancements are not stifled [22]. Machines that are powered by artificial intelligence present interesting new challenges to deal with. For instance, Volkswagen engines were found guilty of emissions testing fraud. While in reality, the engines spewed pollutants up to 40 times above regulatory norms in the United States, they were clever enough to detect the emission test and automatically control and lower their output. Similarly, Apple Inc. was accused of slowing down its old iPhone. This entails determining what value system the robots will employ, particularly in cases of ethical quandaries, such as selecting one of two terrible decisions. Who will own the machines after they have been put to commercial use? Will it be the manufacturer, the owner, the community, or other stakeholders? In addition, there will likely be conflict when these different parties want to gain something from the machines. Asimov's law of robotics will further create challenges when machines start to have consciousness. There is a possibility that a simulation of AI may produce conscious machines, but these conscious machines may not share many of our human values and goals. It is extremely difficult for a robot to comprehend the whole breadth of human language and the experiences it reflects. In different circumstances, broad behavioural goals like preventing human injury or ensuring the existence of a robot can mean different things. The following are Asimov's three laws: First, a robot may not harm a human or allow a human to be harmed by its inactivity. Second, robots must conform to human commands unless such directives conflict with the First Law. Third, a robot must defend itself as long as it does not violate the First or Second Laws. Asimov later added the fourth law, which states that a robot may not hurt mankind or allow humanity to come into danger by inaction. Although these laws appear to be rational, there have been various arguments raised to illustrate why they are insufficient. These laws, for example, do not apply to ethical dilemmas or instances in which robots must choose between two bad options. A well-known example is a self-driving automobile that must choose between running over pedestrians and sacrificing itself and its occupants. Also, to create an algorithm that will allow a care robot to lend a helping hand to people who have fallen [23]. Since robots are on the verge of becoming our assistants, friends, and co-workers, we must address the more complex scenarios that this will entail, as well as the ethical and safety considerations that this will raise. Because robots will be used in various domains to protect the interests of diverse stakeholders, it would not be feasible to code all types of robots with a single universal moral code. The variety of moral rules followed by robots could be exploited as a marketing tool to promote these artificial intelligence devices. In addition to Asimov's rule, AI engineers should research time-tested moral philosophies that have aided in the understanding of human values and the building of a civilised and sophisticated human civilization. How might the ideas of some of the world's greatest moral thinkers be used for the development of better algorithms and more ethical and value-based robots? These moral ideologies are quite helpful in overcoming ethical difficulties and challenges [24, 25]. Some of these ideas are very supportive in assisting humans in making ethical decisions in life, and if they can help humans, they may also be effective in assisting robots. Due to the large variety of moral philosophies,

only the most common views that are beneficial for making ethical decisions will be discussed.

Normative ethics is the creation of moral codes that have a direct impact on human actions, organizations, and lives. For convenience, we can classify some of the popular normative moral philosophies as consequentialist theories and non-consequentialist theories.

1.2.1 Consequentialist theories

To make a decision based on what is best for everyone, one should look at the two most popular consequentialist theories, *utilitarianism*, and *egoism*. One of the most well-known consequentialist theories of utilitarianism is the first in line [26, 27]. Utilitarianism's most famous proponents are Jeremy Bentham and John Stuart Mill [28–30]. In every given scenario, Bentham views all activities that maximise pleasure and limit misery as being morally moral. The philosophy is best summarised by Bentham's famous remark, "the greatest happiness for the greatest number of people" [31]. The theory is more consequentialist because it focuses on the implications of a specific action on the greater population. Similarly, John Stuart Mill argues that an action is right if it creates the greatest amount of utility for all those who are impacted by it [32, 33]. Based on utilitarianism, AI robots could be trained to make judgments that deliver the greatest happiness to the greatest number of people. In circumstances where they must choose between two separate vices, AI vehicles programmed with utilitarianism will assess the scenario and seek to save the greatest number of human lives possible. Critics may argue, however, that what if an innocent person walking alone on the sidewalk while following all the rules was run over by a utilitarian programmed car full of passengers? This would be a violation of an individual's human rights. However, no one can disagree that the car was following at least an established ethical principle of utilitarianism by picking the best possible alternative to minimise the loss of human life. A human could have done the same thing.

The second important consequentialist theory is the theory of *egoism* by Henry Sidgwick. This theory holds that human conduct should be based exclusively on self-interest [34]. According to Baier, "Egoism recommends that we work for ourselves in the light of our knowledge, or predilections, preferences, likes, and dislikes." [35]. The talisman followed by ethical egoists is "*do well by doing good.*" "Do well, by doing good" is the corporate social responsibility (CSR) business formulae mainly based upon the principle of egoism, wherein the company tries to do good things in the society as part of CSR hoping that society will notice it and will reward by buying more of the company's product. Thus, the company bottom line will improve along with its image. Here, the motive behind doing good is to get rewarded financially. Since the machines are identified with their owners, the self-interest here could be inferred as the interest of the owner or customers. As an enlightened egoist, AI machines could be programmed to do good in the interests of their owners. According to a study done by the Charities Aid Foundation [36], more than half of British adults are more likely to buy a product or use a service from a firm that donates to charity organisations. Thus, AI machines programmed with the egoism principle would do good not because it is a good act, but because it will benefit their owner's self-interest.

1.2.2 Non-consequentialist theories

Consequentialists believe in doing what is best for the majority group's welfare, but non-consequentialists do not believe in infringing on an individual's rights even if it is likely to benefit a larger number of individuals. In contrast to consequentialism, non-consequentialists believe in making decisions based on universally accepted ethical values such as fairness, rights, truth, justice, and commitment, among others. Non-consequentialists think that certain moral principles are inherently valid and morally binding, and it is therefore expected for one to do what is right, even though it may impair the majority group's interests. Non-consequentialist ethics is also known as duty-based ethics since it is based on a duty-based approach and respect for the individual's rights. Moral Absolutism, Moral Nihilism, Immanuel Kant's Categorical Imperative (CI), and John Rawls theory of justice, are notable non-consequentialist ideas. To programme the ethical codes of AI machines, these lovely normative non-consequentialism principles may be simplified and made machine compliant.

Kant believes that religion and morality should be kept separate because if we look at different religions for morality, we will all come up with different solutions [26, 37–39]. The categorical imperative principles of Immanuel Kant are the moral requirements that come from pure reason, and a person must fulfil them [40–44]. Three formulations are most helpful in understanding the CI [45].

The universalization principle was introduced in the first formulation. " Act exclusively in accordance with the maxim that you would like to see become a universal law…", Kant states [45]. This phrase is very similar to the golden rule, which states, "Do unto others as you would have them do unto you," and is found in almost all religious texts [46].

The second formulation is often known as the "humanitarian formula." It focuses on how people should be treated. *Act in such a way that humanity is always an end and never a means, whether in your own person or in the person of another*, Kant states. Humans should not be exploited as tools to achieve a goal, but rather as a goal in the end of themselves, because they are logical and self-aware. For

example, a person who uses a car for everyday trips will cease using it if the car becomes troublesome. However, a person can utilise objects like cars as a mere means, but not human beings, because humans are logical beings and an end in themselves. Humans exist for themselves, unlike automobiles, which are designed to transport passengers. It would be wrong for an AI system to treat humans as "mere means," because humans are reasoning beings who are ends in themselves.

The third formulation states, "*Therefore, every rational being must act as if he were always a legislating member of the universal kingdom of ends through his maxim*." Kant proposes that whatever law any rational creature is subjected to, he (as an end in himself) must be able to view himself as legislating (enacting) it universally. As a result, one should avoid acting on generally applicable maxims. Individual rights, Kant maintains, cannot be infringed upon, even if it is necessary to defend the interests and rights of the majority of people.

As a result, robots based on Kant's principle would prefer individual rights over group rights if they conflict. Immanuel Kant also proposed another normative theory called the "Hypothetical Imperative." Hypothetical imperatives are the commands of the reason that are predicated on "if," then assertion. These are the commands that should be obeyed if one wishes to obtain something [47–51]. The emphasis here is on prudence rather than morality. For example, if you want to make money, you should go to work, or take a break if you're fatigued, etc.

Another non-consequentialist perspective is *moral absolutism*, which holds that any conduct should be judged according to universal moral norms. However, critics of this theory criticise the idea that moral standards are universal and absolute because people from different countries hold differing perspectives. *Moral relativism*, on the other hand, opposes moral absolutism by denying that moral principles are absolute and universal. Moral relativism asserts that morality is a relative idea and that no universal moral standards can be applied to all individuals at all times. It depends on one's cultural background. *Moral nihilism*, also known as ethical nihilism, asserts that there are no objective moral facts, i.e., nothing fundamentally moral or immoral, good or evil, right or wrong because there are no absolute moral truths. Bribery, for example, is not only not bad, but it is also not right, according to a moral nihilist.

Justice, according to John Rawl, is a fair distribution based on a fair procedure, rather than natural law or logic [52]. Artificial intelligence could contribute to the expression of Rawls' theory and the concept of fair justice. John Rawl posed a hypothetical question in 1971: "What would happen if the representatives of society who are responsible for drafting the laws that govern society were unaware of their position in society?" [53]. Rawls refers to this as an "original position," in which lawmakers would be blinded by a veil of ignorance. That is, they will be totally aware of all the basic and uncontroversial truths about science and society, but they will be totally unaware of their place in society. According to Rawls, in such a setting, the laws enacted by legislators will be reasonable and fair since they will avoid bias. They'll endeavour to come up with regulations that are ethical, fair, and equitable for everyone, and that doesn't unfairly favour or disfavour any particular group. In the physical world, such an idea can only be envisioned and is difficult to transfer into reality, but AI shows promise in delivering Rawls' original position concept. The AI machine has enormous potential to represent Rawls' magnificent notion of justice. Because these machines will be free of some of the most basic vices shared by all humans, such as lust, wrath, greed, and attachment, they will create beautiful and just laws for individuals from all walks of life.

1.3 Feminist care ethics and artificial intelligence

It's worth mentioning here that feminist care ethics is a theory that contradicts prominent philosophers' traditional normative ethics. Traditional normative theories and moral philosophies are accused of being gender biassed and disregarding women's moral concerns by the advocates of feminist ethics [54, 55]. The main proponents of feminist care ethics are Carol Gilligan and Nel Noddings. According to Gilligan, men and women look at ethical situations through distinct lenses and viewpoints [56]. Men's ethics is morality based on abstract ideas such as rights and fairness, even at the expense of people's well-being, whereas women's ethics is morality based on caring [57]. The ethics of care differs from traditional ethics in that it believes that women place a greater emphasis on the relationships between the people involved and views morality as a tool for caring for those in relationships and nurturing strong bonds. Men, on the other hand, see ethics as a set of abstract laws that must be followed and applied consistently. Because of their ties to their mothers, women are said to have a higher prevalence of care-based morality. Traditional normative ethics and feminist care ethics vary in the belief that, in traditional normative ethics, moral actors are logical and detached, whereas feminist ethics argues that moral agents are emotional and attached. Traditional ethics presumes that the moral agent is unbiased and suspects his partiality, whereas feminist ethics suspects the moral actor's impartiality since he is emotional and linked. Traditional ethics emphasises universal ethical principles such as rights and justice, whereas feminist ethics emphasises the preservation of ties through virtues such as compassion, caring, understanding, and so on, wherever possible without jeopardising one's own integrity. Traditional normative ethics is mostly an extension of work historically associated with males, whereas feminist ethics is primarily

an extension of work typically connected with women [58]. Acting justly under a care-based morality implies avoiding violence and assisting people in need. While Gilligan believed that care ethics differed from traditional normative ethics, she never declared feminist ethics superior to traditional normative ethics. An AI robot with care ethics could provide excellent care for humans, whether as a nurse, doctor, or someone who assists with housework and errands.

1.4 Human rights and artificial intelligence

Human rights are a notion that allows individuals and groups to have an equal voice and influence over the actions of those in positions of authority. It compares individual strength to that of states and organizations. AI proponents foresee better and more equitable outcomes in the future. By relying on data, the weaknesses associated with human decision-making can be removed [59]. Human existence is made easier by AI, which improves access to healthcare, education, communication, science and technology, agriculture, and government services, among other things. AI technology is saving lives by assisting people with the prediction of natural disasters, the conservation of animals, and the mitigation of the negative effects of climate change. However, AI poses a number of challenges, many of which are related to human rights issues, and AI-powered technology has the potential to cause the very human rights violations and other issues that it is intended to prevent, and even if it is used with the finest intentions, artificial intelligence poses distinct hazards to human rights. Both the United Nations General Assembly and the United Nations Human Rights Council (HRC) passed resolutions in 2019 calling for the application of international human rights law to AI and other emerging digital technologies, with the General Assembly warning that "profiling, automated decision-making, and machine learning technologies… without proper safeguards, may lead to decisions that have the potential to affect the human rights of others."

AI has spawned new forms of oppression that disproportionately impact the most disadvantaged and weak. "Automation may help exacerbate the widening gap in the global economy. Due to the various platforms that AI enables, small groups of people are now able to make big profits while employing a limited number of others. This is unavoidable; growth is inevitable; but, it is also socially damaging," writes Stephen Hawking [60]. AI may have an adverse impact on employment. If huge numbers of individuals are unable to obtain employment, many of them will struggle to support themselves and their families. Researchers are investigating the means of maintaining a level of life in the face of the unpredictability of the work market. One example is the provision of a universal basic income for all people.

Credit scores and loan report screening were built using algorithms and have been in use for a long time. Machine learning systems use machine learning to assess non-financial data points to arrive at creditworthiness, such as where someone lives, what they do on the internet, and what they buy. E-scores are the result of these systems. These scores may be discriminatory towards the marginalised and could lead to financial discrimination. Furthermore, AI technology, such as facial recognition, may produce errors if the user has dark skin. This violates the idea of equal rights and equal opportunity.

Because many prejudices are reinforced through the use of AI in the criminal justice system, AI has the potential to pose a threat to the criminal justice process. For example, AI might help with "risk scoring" and "predictive policing." While "risk scoring" aids in identifying whether a defendant is likely to re-commit a crime, 'predictive policing' aids in predicting crime by combining insights from numerous data sets. The use of machine learning for ‘risk scoring’ of defendants may be useful in eliminating the recognised human bias of the judges in granting bail and awarding sentences, whereas predictive policing may be useful to effectively utilise often-limited police resources to prevent crime. However, experts have raised the apprehension that, instead of facilitating, the AI systems' recommendations may instead aggravate problems, either directly or indirectly, by including elements that are substitutes for bias.

Because of the vast amount of data humans produce due to the use of the internet and IoT devices, the threats associated with AI's ability to track and analyse our digital lives are amplified. AI could be used to collect and analyse all of this data for a variety of purposes, including micro-targeted advertising, public transportation optimization, and government surveillance of civilians. Not only are there significant hazards to privacy in such a world, but it also begs the question of whether data protection is really conceivable. This is a breach of the right to privacy and protection.

Drone technology, GPS technology, fingerprint recognition, face detection, retina recognition, and other technologies could be utilised for predictive policing and surveillance of people's movements. These technologies could be utilised for policing and tracking people's movements. Even if done for justifiable public safety concerns, this may risk infringing on one's right to freedom of movement [61].

AI has the power to influence public debate and fabricate and promote content that causes strife, such as conflict, discrimination, animosity, or violence. This goes against Article 20 of the International Covenant on Civil and Political Rights (ICCPR), which states that any propaganda for war should be illegal and incitement to discrimination, animosity, or violence is explicitly forbidden by the law.

Additionally, there have been concerns that if AI-assisted health and reproductive screening is utilised, and individuals

discover themselves unlikely to conceive, they could restrict themselves from getting married or marrying a specific individual. AI-powered DNA and genetics testing have the potential to be used in projects to ensure that only desirable traits are passed on to future generations [62]. This goes against Article 23 of the ICCPR, which is related to the right to marry, children's rights, and family rights. These criticisms notwithstanding, many are concerned that AI-supported tracking and prediction of student performance may restrict access to various educational options and so undermine a student's right to education. This approach would dismiss students who overcome obstacles to gain success in school and the workforce and instead promote the disparity that already exists. This will thus be a violation of Article 13 of the ICCPR, which has to do with universal education.

Machines that enable governments to locate and crush cultural groupings can cause people to cease all cultural activities, both at an individual and communal level. Criminalizing particular cultures could be possible if AI is deployed. Artificial intelligence-powered surveillance could also be used to limit and prevent political participation, such as by recognising and demotivating specific sections of the population from casting ballots.

2 Conclusion

Now that computers are doing all of the work that humans do and are gaining consciousness, the prospect of applying ethical codes of conduct to them in the same way that humans do will have to be investigated. In this case, normative theories can be highly useful because they are secular and do not require theological or philosophical understanding to comprehend. Because these ideas were proposed by well-known philosophers, their validity is self-evident. If programmers could somehow encode these fundamental normative rules into a computer algorithm, it would be possible to design ethical AI robots. It is important to note that it will not be appropriate to code all of the machines using the same theory or principle, since each machine will do distinct tasks, necessitating diversity in their decisions. However, the notion of least harm should always be universal and a component of their decision-making process. These normative theories are simple to comprehend and use, and they do not necessitate a thorough knowledge of challenging philosophical or religious concepts. These beliefs assert that good and evil can be discovered solely via reasoning and that any logical conclusion may be reached without a complete understanding of philosophy or religion. Some of these concepts are quite useful in guiding humans to make ethical choices. Non-consequentialists think that judgments should be made based on widely acknowledged ethical ideals, including fairness, rights, truth, justice, and commitment, whereas consequentialist ideas include "doing well by doing good" and "self-interest." Feminist ethics is a paradigm that runs counter to the established normative ethics of notable thinkers. Feminist ethics supporters criticise traditional normative theories and moral philosophies for being gender-biassed and dismiss them.

If individual rights and group rights conflict, AI would prioritise individual rights over group rights. It is suggested that.

- It would be impractical to code all AI robots functioning in various fields with a single universal code of ethics.
- The notion of causing the least amount of harm to humans should be at the heart of AI robot ethics.
- At least one or more feasible normative theories should be followed by AI devices.
- As a result of the variation in coding moral standards, robots' ethical decision-making will be more diverse. It might also be an AI robot's USP (unique selling proposition).
- For AI machines, there should be some kind of standardisation and protocol, and each computer should adhere to it.

In the field of human rights, scholars have voiced concerns that AI-powered technology can produce the very violations of human rights and other issues that it is supposed to avoid. Experts are concerned that, rather than helping, AI systems may exacerbate human rights difficulties, either directly or indirectly, by integrating aspects that act as biased substitutes. However, though there are concerns about AI's impact on human rights, it is reasonable to conclude that, while AI may not be the panacea for the ailment of poor decision-making, it should not be a cause to dismiss it outright.

Funding No funding has been received for this manuscript.

Declarations

Conflict of interest There is no conflict of interest.

References

1. Lloyd, D.: Frankenstein's children: artificial intelligence and human value. Metaphilosophy **16**(4), 307–318 (1985)
2. Gills, A.S.: Internet of Things (IoT). TechTarget. https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT (2020). Accessed 21 June 2021.
3. Poudel, S.: Internet of things: underlying technologies, interoperability, and threats to privacy and security. Berkeley Technol. Law J. **31**(2), 997–1022 (2016)

4. Miller, S.: Artificial intelligence and its impact on legal technology: to boldly go where no legal department has gone before Thompson Reuters. https://legal.thomsonreuters.com/en/insights/articles/ai-and-its-impact-on-legal-technology (2017). Accessed 20 June 2021
5. Kumar, S., Choudhury, S.: Ancient vedic literature and human rights: resonances and dissonances. Cogent Soc. Sci. **7**(1), 1858562 (2021)
6. Kumar, S., Rai, A.: Business ethics. Cengage Learning, New Delhi (2019)
7. Bernstein, J.: Profiles: A.I. The New Yorker. https://www.newyorker.com/magazine/1981/12/14/a-i (1981)
8. Dennett, D.: When philosophers encounter artificial intelligence. Daedalus **117**(1), 283–295 (1988)
9. Arata, H., Hale, B.: Smart bases, smart decisions. Cyber Def. Rev. **3**(1), 69–78 (2018)
10. Costigan, S., Lindstrom, G.: Policy and the internet of things. Connections **15**(2), 9–18 (2016)
11. Miller, M.: The Internet of Things: How Smart TVs, Smart Cars, Smart Homes, and Smart Cities are Changing the World. Indianapolis, Que (2015)
12. Boden, M.A.: The philosophy of artificial intelligence. Oxford University Press, Oxford (1990)
13. Buchholz, S., Mariani, J., Routh, A.,Keyal, A. & Kishnani, P.: The Realist's guide to quantum technology and national security. Deloitte's Center for government Insights. https://www2.deloitte.com/us/en/insights/industry/public-sector/the-impact-of-quantum-technology-on-national-security.html (2020)
14. Newell, A., Simon, H.A.: Human Problem Solving. Echo Point Books & Media, Brattleboro (2019)
15. Crevier, D.: AI: the tumultuous search for artificial intelligence. Basic Books, New York (1993)
16. Searle, J.: Minds, brains and programs. Behav. Brain Sci. **3**(3), 417–457 (1980). https://doi.org/10.1017/S0140525X00005756
17. Guo, T.: Alan Turing: Artificial intelligence as human self-knowledge. Anthropol. Today **31**(6), 3–7 (2015)
18. Shelley, M.W.: Frankenstein, or the Modern Prometheus. Bibliologica Press, Unley, SA (2021)
19. Geraci, R.M.: Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality. Oxford University Press, New York (2012)
20. Ortega-Rodriguez, M., Solís-Sánchez, H.: Is there a relationship between Stephen Hawking's worldview and his physical disability? On the importance of cognitive diversity. Cornell University. Preprint at arXiv:1804.09012 [physics.pop-ph] (2018)
21. Hawking, S.: Personal Interview to Rory Cellan-Jones, BBC. https://www.bbc.com/news/technology-30290540 (2014). Accessed 20 June 2021.
22. Perc, M., Ozer, M., Hojnik, J.: Social and juristic challenges of artificial intelligence. Palgrave Commun. **5**, 61 (2019). https://doi.org/10.1057/s41599-019-0278-x
23. Schmarzo, W., Vantara, H.: Asimov's 4th Law of Robotics. KDnuggets. https://www.kdnuggets.com/2017/09/asimov-4th-law-robotics.html#:~:text=A%20robot%20may%20not%20injure,being%20to%20come%20to%20harm.&text=A%20robot%20must%20protect%20its,the%20first%20or%20second%20laws (2017). Accessed 20 June 2021.
24. Hasnas, J.: The normative theories of business ethics: a guide for the perplexed. Bus. Ethics Q. **8**, 19–42 (1998)
25. Heath, J.: Business ethics without stakeholders. Bus. Ethics Q. **16**(4), 533 (2006)
26. Hartman, L.P.: Perspectives in business ethics. McGraw-Hill, Boston (2002)
27. Trevino, L.K., Nelson, K.A.: Managing Business Ethics: Straight Talk About How To Do It Right, p. 40. Wiley, Hoboken (2011)
28. Bowie, N.E., Schneider, M.E.: Business ethics for dummies. Wiley, Hoboken (2011)
29. Ferrell, O.C., Fraedrich, J., Ferrell, L.: Business Ethics. Cengage, Mason, OH (2018)
30. Painter-Moreland, M.: Business Ethics as Practice, p. 53. Cambridge University Press, Cambridge (2010)
31. Bentham, J., Mill, J.S., Mill, J.S.: The Utilitarians: An introduction to the principles of morals and legislation [by] Jeremy Bentham. Utilitarianism, and On liberty [by] John Stuart Mill. Doubleday, Garden City, NY (1961)
32. Mill, J.S.: Utilitarianism. GRIN Verlag GmbH, München (2008)
33. Mill, J.S., Lazari-Radek, K., Singer, P.: Utilitarianism. W.W. Norton & Company, New York (2022)
34. Sidgwick, H.: The Methods of Ethics, p. 119. Macmillan Publishers, London (1907)
35. Baier, K.: Moral obligations. Am. Philos. Q. **3**, 210–226 (1966)
36. Johnston, C.: Why do corporates give? Raconteur. https://www.raconteur.net/why-do-corporates-give/ (2014). Accessed 25 June 2021.
37. Allison, H.E.: Kant's Groundwork for the metaphysics of morals: A commentary. Oxford University Press, Oxford (2011)
38. Burch, M.: Normativity, meaning, and the promise of phenomenology. Routledge, New York (2019)
39. Wood, A.W.: Kant and Religion. Cambridge University Press, Cambridge (2020)
40. Kant, I., Korsgaard, C.: Frontmatter. In: Gregor, M. (ed.) Kant: Groundwork of the Metaphysics of Morals. Cambridge Texts in the History of Philosophy, p. I–IV. Cambridge University Press, Cambridge (1998)
41. Stern, R.: Kant. In: Understanding Moral Obligation: Kant, Hegel, Kierkegaard (Modern European Philosophy), pp. 5–6. Cambridge University Press, Cambridge (2011)
42. Timmermann, J.: Kant. In: Golob, S., Timmermann, J. (eds.) The Cambridge History of Moral Philosophy, pp. 394–409. Cambridge University Press, Cambridge (2017)
43. Timmons, M.: The philosophical and practical significance of Kant's Universality formulations of the categorical imperative. Ann. Rev. Law Ethics **13**, 313–333 (2005)
44. Watkins, E.: Kant on Laws. Cambridge University Press, Cambridge (2019)
45. Kant, I.: "First Section: Transition from the Common Rational Knowledge of Morals to the Philosophical" Groundwork of the Metaphysics of the Morals. Harper and Row Publishers, New York (1997)
46. Hoche, H., Knoop, M.: Logical relations between Kant's categorical imperative and the two golden rules. Ann. Rev. Law Ethics **18**, 483–518 (2010)
47. Kitcher, P.: Kant's argument for the categorical imperative. Noûs **38**(4), 555–584 (2004)
48. Marshall, J.: Hypothetical imperatives. Am. Philos. Q. **19**(1), 105–114 (1982)
49. Morrisson, I.: On Kantian Maxims: a reconciliation of the incorporation thesis and weakness of the Will. Hist. Philos. Q. **22**(1), 73–89 (2005)
50. Shand, J.: Kant, respect, and hypothetical acts. Philosophy **90**(353), 505–518 (2015)
51. Shaver, R.: Korsgaard on hypothetical imperatives. Philos. Stud. **129**(2), 335–347 (2006)
52. Rawl, J.: A Theory of Justice. Harvard University Press, Cambridge, MA (1971)
53. Gorovitz, S.: John Rawls: a theory of justice. In: de Crespigny, A., Minogue, K. (eds.) Contemporary Political Philosophers, pp. 272–289. Dodd, Mead and Co, New York (1975)
54. Hamington, M., Sander-Staudt, M.: Applying Care Ethics to Business. Springer, Dordrecht (2011)

55. Held, V.: The Ethics of Care: Personal, Political, and Global. Oxford University Press, New York (2007)
56. Gilligan, C.: In a DIFFERENT VOICE: PSYCHOLOGICAL THEORY and Women's Development. Harvard University Press, Cambridge (1982)
57. Noddings, N.: Caring: A Feminine Approach to Ethics & Moral Education. Univ. of Calif. Press, Berkeley (1986)
58. Cavanagh, G.F., Moberg, D.J., Velasquez, M.: Making business ethics practical. Bus. Ethics Q. **5**(3), 402 (1995)
59. Santow, E.: Can artificial intelligence be trusted with our human rights? Aust. Q. **91**(4), 10–17 (2020)
60. Hawking, S.: This is the most dangerous time for our planet. UK: Guardian. https://www.theguardian.com/commentisfree/2016/dec/01/stephen-hawking-dangerous-time-planet-inequality (2016). Accessed 21 June 2021.
61. Kumar, S., Choudhury, S.: Migrant Workers and Human Rights: A Critical Study on India's COVID-19 Lockdown Policy. Elsevier Social Sciences and Humanities. Elsevier, Amsterdam (2021)
62. Access Now: Human Rights in the age of artificial intelligence. https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf (2018). Accessed 20 June 2021