Proposed Way to Inculcate Morality in Artificial Intelligence

Richa Kapoor Mehra
O. P. Jindal Global University, Sonipat, India

Keywords: Artificial Intelligence, Ethical Theories, Utilitarianism, Categorical Imperative, Virtue Ethics.

Abstract: Recently one of the most trending, memorable and unique news hitting the social media that showcased one of the best uses of technology is in the form of the clad draped in a traditional saree to accept loan documents from bank branch. Robotic company ASIMOV sent its saree- clad advanced robot SAYABOT to receive loan documents from the bank. One of the main reasons for the popularity of such humanoids is due to their capability of performing task in almost all sectors such as education, healthcare, hospitality and banking. In the 21st century, human dependency on robots and other gadgets has increased drastically, as a result giving rise to plenty of challenges. With AI becoming smarter day by day in taking decisions and getting involved in other activities results into various ethical concerns. Few questions that one can ponder upon- can these digitally empowered Bots who behave like humans and dress like humans, act in socially and morally responsible way? Can Artificial Intelligence be said to be moral agents? Can artificial intelligence be said to have the ability to discern right from wrong? Who bears responsibility of wrong doings by Artificial Intelligence? Can we hold artificial intelligence accountable for its own actions? Is there a way to equip artificial intelligent as artificial moral agents? This is where role of ethics comes in. The increasing use of AI in 21st century in more efficient and faster way poses various challenges on society and its beings. Now one of the significant question is what should be done to ensure that AI outbreak gives rise to positive results which benefit the society as a whole? This article is an attempt to answer the above-mentioned questions by showcasing ethics as a means to tame and instruct artificial intelligence to ensure they act in a morally responsible way. The aim of this article is to determine which ethical theory is best suitable theory that would provide results in favor of humanity. The first section of this article deals with consequentialist theory of utilitarianism. In the second section an attempt has been made to comprehend how compatible artificial intelligence would be with Aristotle's virtue theory. The third and last section showcase the compatibility of artificial intelligence with deontology theory of Immanuel Kant.

1 INTRODUCTION

Since time immemorial, morality is considered to be the very essence of human beings. Morality is the fundamental human trait that makes us responsible moral agents. Human dignity and moralresponsibility are the key essence of human traits. Humans are moral agents; they have the abilityto discern from right and wrong actions and are considered accountable for their actions. With thegrowing use and demand of artificial intelligence, it is imperative to check the actions of AI and ensure that their actions are ethically governed. Before we move forward, it is imperative to first know what is meant by artificial intelligence.

AI can be understood as intelligence that is artificially created and that is demonstrated by a machine. From Alexa to touch screen dispensers and other gadgets, we can see our dependence onmany such intelligent items. The term AI can be understood as intelligent agent that is designed and created by human beings; is so intelligent that it perceives its environment and becomes smarter. The most common existing form if artificial intelligence is simple AI, these are machines that rely on the preprogrammed data or set of rules hence may not be very efficient decision makers. On the other hand there is *Machine Learning*, which permits machines to acquire from the surroundings without being overtly automated thus equip themselves as better decision makers. This type of technology is often referred as complex artificial intelligence; they have better decision-making capabilities as they get involved in integrating and scrutinizing data relating to a specific task. Since these complex AI takes

Mehra, R.
Proposed Way to Inculcate Morality in Artificial Intelligence.
DOI: 10.5220/0012615400003739
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 1st International Conference on Artificial Intelligence for Internet of Things: Accelerating Innovation in Industry and Consumer Electronics (AI4IoT 2023)*, pages 319-321
ISBN: 978-989-758-661-3
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

decision on the bases of pre given data, so to ensure they act in morally appropriate way; I suggest to equip them with certain ethical theories. Thus, to ensure best results we have to feed ethical data to the AI so that they act in ethically smarter and better way. Now the question that can be posed- how one become aware of ethical data? To unfold this question, we will explore various ethical theories and will discover which theory yield the best outcome for humanity?

Ethical theories provide us with the foundation for decision making. Ethical theories help in distinguishing between right and wrong actions. These theories help us in getting the unified account of our moral obligations. If AI programmers instruct AI with the most suitable ethical theories, then such theories will benefit AI in taking morally right decisions. To determine which ethical theory would yield the best result it is imperative to unpack few most relevant ethical theories.

2 CONSEQUENTIALIST THEORY

The judgement of action as good or bad is determined on the basis of the outcome of the action. An action giving rise to good consequences is considered as good action, on the other hand if the consequences of an action are bad; the action is considered as bad. The morality of action in consequentialist theory is determined on the basis of its outcome. This theory judges an action to be good or bad on the basis of its consequences. For instance, most of people would believe that lying is bad; but for consequentialist, if a lie saves someone's life then it is not bad to lie. There are primarily two types of consequentialists- utilitarianists and hedonists. One of the famous forms of consequentialism is utilitarianism, it aims to perform those actions which may lead to maximum benefits to maximum number of people. In other words, an action that brings more benefit than harm is considered as good action, and the action that produces more harm than good is consideredas bad action and should be avoided. Two famous English thinkers who have defended utilitarianism include Jeremy Bentham and John Stuart Mill. Utilitarian's doctrine can be expressed in the form of a single principle, the greatest happiness principle: the rightness of an action is determined by its contribution to the happiness of everyone affected by it. On the other hand, hedonists believe that an action is good if it results into pleasure and it's bad if it gives rise to pain. To achieve the maxim of maximum happiness for maximum number of people, for utilitarians it is imperative to increase number of good things in the world and to decrease the number of bad things.

3 VIRTUE THEORY

This theory has gained popularity in past 20 years, virtue theory is contrary to consequentialist theory. Aristotle, the famous ancient Greek philosopher advocated this theory. The prime emphasis of this theory is development of one's character, it focused on the development of human qualities. Prime questions emphasized by this theory includes - "How should a person live?" and "What is the good life?" and "What are values?"

Aristotle believed that virtues can be imbibed through habit, according to him by practicing virtues repeatedly a person develops moral character. Through continuous practice of being truthful, just, generous, compassionate and so on, a person develops a decent ethical character. This results into making correct choices specially during ethical dilemma situations.

Aristotle then observes that a thing is called good when it performsits function well same is applicable to human beings; for him when a person channelize reason and take decision based on reasoned inquiry, they are believed to perform well. For Aristotle, man's highest function is to reason well and a good man is one who reason's well. However, virtue theory may not be said to be compatible with artificial intelligence since having virtuoustraits in artificial intelligence does not ensure that the decision making will always be in favor of other human beings. In the up-coming section deontologicaltheory has been deliberated to explore the compatibility of this theory with the artificial intelligence.

4 DEONTOLOGICAL THEORY

In contrast to consequentialist theory, deontology theory was proposed by Immanuel Kant, it contemplates actions to be ethically right or wrong in and of themselves, irrespective of their outcome; they use universal rules to distinguish right from wrong. The ethical theory of categorical imperative is core ethical theory of Kant. The term categorical imperative, basically means "absolute command." According to Kant, there is only one categorical

imperative, which he presents in three diverse principles:

1: The law of universalizability: this principle states that an action is considered to be right if it can be universalized. For instance, breaking promise or telling a lie; before performing these actions, a person has to see whether it can be universalized. Kant puts it this way: "Act as though the maxim of your action were by your will to become a universal law of nature.

2: The Principle of Ends states that: "So act as to treat humanity, whether in your own person or in that of any other, inevery case as an end and never as merely a means." To put it differently, this law states that we should treat human beings as end in themselves and never use them as a means to fulfill some other end., To put it differently, individuals should give respect to each human being and should not use them as a commodity to accomplish the desired results. Kant states that since we humans have rational capacity or ability to reason we can make use to reason to perform right actions.

3: The Principle of Autonomy this principle states that we are free rational beings and due to this pre-given rationality, we are able to determine the differences between right and wrong actions. To differentiate between right and wrong actions, we do not have dependent upon others. However, for Kant the onus lies on humans to discover and identify differences between morally right and wrong actions. We have to use our ability to reason to assist us in implementing the categorical imperative and make our own decisions, rather than relying on someone else to tell us what to do. Kant puts it this way: "So act that your will can regard itself atthe same time as making universal law through its maxims.

Now putting these ethical theories in context of artificial intelligence decision- making, the differences in the theories would result in the difference in decision making. For instance, a consequentialist AI might think that killing one evil human being would be a just act if it results into saving thousands of other human beings. However contrary to the normative approach, a virtue theory AI emphasizes on development of moral traits of an individual. AI following deontological theory unlike the consequentialist theory would regard the act of killing as a wrong act.

5 CONCLUSION

With the advancement of technology and digital revolution in place, it is believed AI to have rational thinking capacity if it involves in a pre-given design of logical thinking basis which it vindicates and perform actions. If applied to AI, deontology theory is most appropriate theory because of its ordering actions and categorizing it on the basis of rational dimensions of the other moral agents. It uses universal law to distinguish right from wrong. For Kant, humans are distinctly superior to other beings due to its rational capacities. This status would also apply to artificial intelligence. In my opinion, out of the above mentioned ethical theories, AI must be automated with deontological. Since AI have the potential to make incredibly complex moral decisions, it is important that humans are able to identify the logic used in a given decision in a transparent way, so as to accurately determine the morality of the action in question.

REFERENCES

Abelson, Raziel and Kai Nielsen. "Ethics, History of" in *Encyclopedia of Philosophy*. Ed. DonaldM. Borchert, 394-439, 2006.

Anderson, M., Anderson, S; *Machine ethics: creating an ethical intelligent agent*. AI Mag. 28(4),2007.

Bentham, J. *Introduction to the principles of morals and legislation*. Blackwell's political texts. Blackwell, Oxford, 1789. Intr. de W. Harrison, 1967.

Bernard Williams. Negative responsibility: and two examples. *Utilitarianism: For and against*,pages 97-118, 1973.

Colin Allen, Gary Varner, and Jason Zinser. Prolegomena to any future artificial moral agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 12(3):251-261, 2000.

Driver, J. The history of utilitarianism. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta,Ed., summer 2009.

Lafollette, Hugh, ed. *The Blackwell Guide to Ethical Theory*. Malden: Blackwell Publishers Inc.,2000.

Moore. J. H; *The nature, importance, and difficulty of machine ethics*. IEEE Intell. Syst. 21(4). 18-21 (2006).

S. Matthew Liao, A Short Introduction to the Ethics of Artificial Intelligence In: *Ethics of ArtificialIntelligence*. Edited by: S. Matthew Liao, Oxford University Press, 2020.