IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Ethics of Machine: How to make Ethical Decision

Arjun Kumar
Ph.D Research scholar
University of Delhi

Abstract:

The purpose of this paper is to explore an answer to the question whether it is necessary to artificially construct moral algorithm in order to act morally good or bad.On the basis of moral algorithms AI (machine, robots, etc) make moral decision in the various situation of human life. Firstly,the paper introduces the definition of artificial intelligence and how it works .Further the paper describes specific theory of morality; how it is follow by humans to make decision. Next, the paper narrates some moral algorithms that had been created/developed by engineers. Next, the paper tells what criteria of moral actions are. Further how do machines make ethical decision on the basis of coded moral algorithm and what can be its consequences. The paper concludes that Machine can not take any ethical decisions in the various situations freely.

Keynotes:

Artificial Intelligence, Utilitarianism, Deontology, Ethical Algorithms, Ethical Decision, Free Will, Decision Making Process.

Introduction:

The highest objective of science of artificial intelligence has been to create artifact that can be regarded as equal or even superior to a human in the case of intelligence. This goal is made yet more complicated in a specific field called ethical decision making of AI. Alan Turing introduced TT (Turing test) in 1950, according to them a computer able to pass TT should be declared a thinking machine. But TT and various other tests tell that machine can only mimic but can not think like humans who can behave intelligently."The fundamental goal [of AI reseach] is not merely to mimic intelligently or produce some clever fake, Not at all. AI wants only the genuine: machine with minds ,in the full and literal sense. This is not science fiction, but real science ,based on a theoretical conception as deep as it is daring ; namely ,we are, at root ,computers ourselves".(Haugeland , *Artificial intelligence*,2.)

In the process of developing AI, we find a lot of machines and artifacts which can learn algorithm and symbolize our activities in machinery language like google, facebook, twitter etc are doing very well .The most significant question for AI in human life is concerning with situations and dilemmas which men face in daily life and make decision according to that. Every action of man is valuable because men have free will, cognition, thinking, capacity, mind and others but machines have mind which are coded in machinery language. Simply we can say them information sand data. In artificial intelligence, all program are running in software formation which was systematically coded. In this era everything can be coded in algorithms even ethical rules and theories.Ethicist have spent many centuries developing ethical theories like Utilitarianism, Deontology to go beyond our intuitive moral judgment in making predictions about which action are good or wrong. Theories aim at both internal consistency and external predictive accuracy. The demand of clarity and

consistency pushes moral judgment to articulate ways of measuring entities like 'property' or 'harm'; it also inspires them to create general rules that can resolve conflicts between violation of property and harm . Principle like "act in a way that maximizes happiness'' a short of extent, we will have to sacrifies our some happiness for the more net good consequences. Other principles,like "don't intentionally do wrong'' probably it may be happen but agent had not that intention to do. Just like our moral grammar or strategies for cooperation games, theories like utilitarianism, deontology can be turned into algorithms but if each theory is internally consistent and naturally incompatible, which one we program into our machine? I propose that moral theory,just like a scientific one, must be more than just internally consistent. It must also make external prediction that can be used to evaluate the theory; the moral theories are rationalization, they are attempted to clarify and generalize our adaptive moral grammar. They can be evaluated by how effectively they solve cooperation problems. Can machine make ethical decision or not? What will be its social, legal consequences?

Ethical theories.

Utilitarianism: actions are wrong whenever their consequences produce more overall suffering for everyone, and permissible when their consequences result in more net happiness. Anyone whose happiness is affected by an action should consider evaluating their action. Consequences are usually measured in terms of probability with more likely happiness counting for more than less likely happiness. Most harmful actions are usually wrong but sometimes it may be acceptable to cause some suffering for generating greater overall happiness. Historical advocates include Jeremy Bentham and John Stuart mill.

Deontology: morality of an action should be based on whether that action itself is right or wrong under a series of rules, rather than based on the consequences of the action, is sometimes described as duty. Action are wrong because they can not be universally applied as a rule to all other people without producing logical inconsistency.

These theory all agree that homicide and to harm humans are morally wrong, but they have different reason for why these action are wrong. Men have moral reasons to treat human being in certain ways and make decision on the basis of that situation.

Implimentation of ethical theories in machine:

The aim, of machine ethics, is to create a machine that itself follows an ideal ethical principles or set of principles. The action of machine is guided by ethical principles. These principles are used to making decision or contemplating about possible courses of action it could take. Men follow these ethical principles to make decision in daily life. Men observe various situations and then react according to that dilemma and situations .Man's decision is not good every times, sometime the consequences of action are bad.

Utilitarian algorithmhas been developed by Michal Anderson, Susan Anderson and ChrisArmen (2005). The utilitarian algorithm uses self-report measurement of intensity of pleasure or pain ,the duration pleasure or pain and likelihood of the event:

"The algorithm is to compute the best action , that which derives the net pleasure , from all alternative actions. It requires as input the number of people affected and for each person ,the intensity of the pleasure/displeasure (for example, on a scale of 2 to -2) , the duration of the pleasure of the pleasure/displeasure or displeasure will occur , for each possible action. For each person, the algorithm computes the product of intensity, the duration, and the probability, to obtain the net pleasure for the person . it then adds the individual net pleasures to obtain the total net pleasure: (Total net pleasure =Σ(intensity x duration x probability) for each affected individual. this computation would be performed for each alternative action. the action with the highest total net pleasure is right action." (Anderson and Anderson,Machine Ethics, 18).

Above, we find that machine can find the possible good action in the alternative actions. Machine computes the intensity, durability and probability of that person who make decision, and gives us a output in form of action which produces good consequences. Here is a plus point for machine whichwill have to be considered.Machines follow the principles strictly, it does not get affected by the emotions or other activity but men always consider these and his action are affected accordingly. If machine acts as an advisor to beings and think as utilitarian point of view, it would prompt to human user to consider alternative actions that might result in greater net good consequences.

Kantian algorithms developed by Thomas Power (2006), where a machine rejects actions whenever they are the result of a contradiction in its background belief, purposes(intentions), and context.

"A rule based ethical theory is a good candidate for the practical reasoning of machine ethics because it generates duties or rules for actions. The rules are computationally tractable. Thomas power considers the three views how to categorical imperative works: mere consistency, commonsense practical reasoningand coherency.''(Powers,*prospects for Kantian machine,47*).

In Kantian algorithm, Thomas Power holds some view and discuss that categorical imperative supplies a procedure for deriving rules: "Act only according to that maxim whereby you can at the same time will that it should become a universal law''

In both ethical theory based algorithm one can take a standpoint by analyzing input principles, computes the situation and give an exact action which does not only have good consequences and can become universal rule also. If we think deeply and go to that point, how do a human being make ethical decision,if one were to contemplate deeply about ethical decision making in human beings one can that our decision reflect some biases and are thereby not neutral.

Criteria of decision making:

There is a long chain to make an ethical decision. If men make decision from a point of view of any principles, there should be some objectives which are consciousness, free will, mind, agency .when men make decision, they consider to some of these objectives. He awares to his activity and he has self-thought , choices , goals,freedom . Moreover men have mind onwhich recognize acts as good or bad .Men have some aims/goals. Depending on these aims/goals, Means are chosen and choices are opted for decision making. This process is taken in every theoretical form even deontology, there are some choices but men make decision by listening their inner voice.

Action of machine (decision making):

Ethical principles are coded in algorithm which are machinery language.Machine make decision on the basis of these algorithm. According to utilitarian, machine can complete durability,probability,intensity and observe pre decisionsthen machine give output which ismore net good consequences. But here is a problem, machines have no free will. Machines are guided by their programmer in the actions. Machine can not think automatically, it needs some inputs. Machine have no mind like humans so machine can think only in a fix dimension. Simply we can say machine are behaving like puppet. But it can be notice a point that machinescan not be biased because their action are totally guided by algorithms. Algorithm can be biased because it is biased of programmer which gets reflected on the program. But when men make ethical decision; they can be partial to his friend, family etc. But machines have no consciousness and can not be partial. It is true that machines can be aware of the society, situation,environment etc,but not like humans.

Hence, Kantian machine can not make a good decision according to situation. Machine can read, learn and take a decision but with some fix rule. Theories aim at both internal consistency and external accuracy. Machine can make decision or give us predictive accuracy in some cases but in human life, there are a lot of

situations and problems.It is not easy for machine to make accurate decision on the base of any particular principle.

Conclusion:

I conclude that it is good that ethical algorithm has been developed and may be working well but there is a problem. That type of machine, only can give advise about issues which have more information and few ideas. It can help us to some situations. Yes, this is true that it depend on the owner that will listen to their view but it is totally depend on the owner that he is accepting that view or not,because men have free will , consciousness , mind ; they can make decision according to the situations . But if we think that machines can freely act and make decision and survive in human society. This may be so dangerous because if machines can behave ethically so it can also behave unethically then the result of decision could be so wrong . Some criteria are commonly proposed to make any ethical decision.

1. The capacity for phenomenal experience or qualia, such as the capacity to feel pain and suffer.

2. A set of capacities associated with higher intelligence, such as self-awareness and being a reason-response agent.

3. Mind with extraordinary qualities.

References:

1. Powers,Thomas."*Prospects for Kantian Machine*."Intelligent Systems, IEEE Vol 21(2006).

2. Anderson,Michael and Susan Anderson."*Machine Ethics: Creating an Ethical Intelligent Agent.*"Ai Magazine vol 28(2007).

3.Mizzoni,John.*Ethics the basics 2ed*.Singapur: John Wiley & Sons,2017.

4. Leben,Derek.*Ethics For Robots*.New York:Routledge,2019.

5. Haugeland,J.*Artificial Intelligence*. Cambridge,MA:MIT Press,1985b.