Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 1 (2024)

A Legal Framework for Determining The Criminal Liability And Punishment For Artificial Intelligence

CS Chaitali Jani [1], Prof. Dr. S.P. Rathor [2]

[1] *Director(I/C) L J School of Law Ahmedabad, Gujarat*

[2] *Director & Professor University School of Law Gujarat University, Ahmedabad, Gujarat*

Abstract:- Today, Artificial Intelligence (AI) is being used increasingly in every field. It has immense positive effects but every coin has two sides. AI is being used in criminal activities as well. Many countries are using AI-powered lethal autonomous weapons. These weapons decide their target and make the decision to kill someone without human intervention. Same way, a self-driven autonomous car may meet with an accident and kill a pedestrian. Who will be liable in both of the above cases? The developer, user, owner, supervisor of AI, or AI itself? For holding a person criminally liable two elements must be present namely, 'actus reus' – a wrongful action and 'mens rea'- a guilty mind or intention for the crime. Can AI have mens rea? Can AI be punished just like a natural person? The article answers such novel issues with plausible arguments. Currently, there are no laws for punishing AI anywhere in the world but this is a dangerous situation where many crimes will go unpunished and it will create an atmosphere of insecurity in the minds of people. To deal with this situation, the article describes three different models that any nation can adopt in its criminal justice system for holding or not holding an AI criminally liable. Even if AI is held criminally liable how to actually punish an AI is a challenging task. The article further explains how the punishment of the death penalty, imprisonment, fine, etc can be extended to an AI. It also explains how the object of punishment can be served by punishing AI by explaining various theories of punishment in the context of AI. A parallel is drawn in the article with corporate criminal liability. Corporates cannot think but still can be held criminally liable but AI can think so why can't AI be held liable for crimes? A critical analysis of punishing AI is done in the article by discussing the pros and cons of punishing AI. The author has done a cost-benefit analysis of punishing AI vis-a-vis other remedies for AI-generated crimes.

Keywords*: Artificial Intelligence (AI), Criminal Liability, Actus Reus, Mens Rea, Negligence, Legal Personhood, Corporate Criminal Liability.*

1. Introduction

Today, Artificial Intelligence (AI) is being used increasingly in different criminal activities. AI is getting active in the dark web, trading drugs, and committing cybercrimes. Many countries have introduced the use of self-driven autonomous cars. In March 2018 a homeless woman named Elaine Herzberg died due to an accident with an Uber test vehicle in Arizona, USA. It is the first case of death in a road accident due to a self-driven car. Many countries are using autonomous weapons in their armed forces. These lethal weapons can decide their target on their own, analyze various modus operandi in a fraction of a second, and can kill people without any human intervention. There are no laws in the criminal justice system of any country yet that can punish these AI-powered autonomous weapons. Today AI is getting involved in our day to day life at breath breath taking rate. Way back in 1981, in Japan, an employee in a Kawasaki motorcycle manufacturing factory was killed by an AI robot working in the same factory. This had happened almost 40 years back. Now, due to rapid changes in technologies, AI's cognitive power has increased to the next level. If AI is not regulated well it can be a threat to society. The question is who should be held accountable for the crimes committed by AI? AI is already

engaging itself in some of the activities that if performed by humans will be considered a crime. Apart from that it is very difficult to trace the crimes committed by AI. The people's apprehension of AI is mainly because AI is still not subject to any criminal law.

2. What Are AI-generated Crimes?

The crimes that are functionally committed by machines and in which there is no identifiable person who has acted with criminal culpability are known as AI-generated crimes. There are cases of AI-related crimes where we cannot hold any natural person liable for that crime. Different types of machines have been used for various activities since ancient times and it has caused harm to humans, animals, properties, and the environment. Most of these were accidents mainly due to negligence of operators or faulty design of the machines. This may draw liability under criminal law, but the liability will be for the users of the machines, operators, owners, or supervisors and not on the machine directly. For example, Today, if someone murders a person with a knife, the person using the knife will be held criminal and not the knife. As a knife is just a tool. However, AI can differ from traditional machines or tools in some critical ways that make the direct application of criminal law more appealing. AI, in particular, can exhibit high degrees of autonomy. In terms of autonomy, AI can make decisions independently without human control. AI can receive inputs from numerous sources in no time, it can set targets, evaluate results against criteria, and adjust behavior as per circumstances to increase its chances of success. If an offense is committed by a person, that person will be criminally liable for it but if the same crime is committed by an AI, without human intervention nobody will be held criminally liable. This situation is very dangerous where crimes are going unpunished in a civilized society.

Today, AI is still not completely autonomous in most cases so AI crimes can be reduced to individuals. For example, If someone develops AI-powered software to steal sensitive data from another person's computer, the software is still a tool and the person who has developed the software will be considered liable, even if the software makes an error and damages the entire computer network system still the person who created it can be held liable, but what if that AI starts making decisions on its own and instead of only stealing the data starts damaging or destroying the entire system or attacks someone on its own.

That time is not far where it is difficult to find any individual responsible for many of the criminal activities as AI is becoming more and more autonomous and complex day by day and can get involved in crimes without any human interference. Also, it is difficult to make one person accountable for AI crime when there are thousands of people involved in the development of such AI for many years. The AI is a result of the hard work and collaboration of many individuals for many years and then after development, the AI may become autonomous and get involved in some crimes then no particular individual can be blamed. Secondly, AI learns from its experience. It collects data from billions of heterogeneous sources. So, an AI developed for a socially beneficial activity may use its data, knowledge, and experience to harm society. In such cases, the developers cannot be held liable. This kind of a situation cannot be predicted or foreseen by any reasonable human being and so, none of the software developers can be held liable criminally. To be criminally liable one must have' mens rea' and 'actus rea' or he may have to be negligent.

As AI is becoming more and more advanced, complex, autonomous, and making independent decisions there will be many such cases. Criminal laws also need to be amended to deal with such situations. Otherwise, these autonomous entities can do any activity, or any crime without being held criminally liable. So, the biggest question is how to deal with this problem. Many such criminal cases where autonomous AI is directly involved are gradually increasing especially in the field of self-driven autonomous cars, dark web, autonomous weapons, etc. These are the areas where immediate regulations for controlling AI are necessary.

3. Models for Criminal Liability of AI

The most important question of criminal law is the question of holding some person (a natural or artificial person like a corporation or AI) criminally liable. For imposing criminal liability on anyone two main elements must exist. The first one is criminal conduct (or omission)-' actus reus' and the other one is criminal intent, mental element- 'mens rea'. If any of this is not present no criminal liability can be imposed on anyone. Many

times, criminal liability is assigned in cases of negligence also when a reasonable person could have easily foreseen and could have avoided by taking reasonable precautions. So, to attract criminal liability the person must have 'mens rea' and 'actus rea' or he may have to be negligent

For example, a child below seven years of age may have killed someone with a loaded gun while playing considering that as his toy gun, the child would not be held criminally liable as mens rea is missing. So, what about AI and its mens rea? Can AI have criminal intent or 'mens rea?' To answer this, three models are discussed here for criminal liability of AI in various situations as suggested by Prof Gabriel Hallevy, an Israeli professor of Criminal Law.

3.1 Artificial Intelligence as A Tool

In this model, AI is considered just a machine, a tool and it is assumed that AI cannot be a perpetrator of a crime, it does not have that mental capacity. So, for any AI-generated crime, there is always a human perpetrator who is to be held responsible. The derivative question is who can be the perpetrator in the case of AI? The perpetrator can be an AI programmer or developer, user or supervisor. The developer might have programmed or designed the AI specifically for committing some crime. Another perpetrator can be the user who might be using the AI for crimes which is developed for other purpose. For example, if An AI designed to follow his user's or owner's command is used by the user or the owner for criminal activity then the real perpetrator to be held criminally liable is his owner or user and not the developer. Same way if the supervisor is negligent or with criminal intentions allows an AI to do an activity that is criminal in nature then the supervisor can be held liable. Here, the actus reus is by AI but mens rea is of a developer, user, owner, or supervisor. AI is just used as an instrument for crime. When an end user makes instrumental use of an innocent agent to commit a crime, the end user is deemed to be a perpetrator. In this model, AI is just like a tool or an animal that is used to commit a crime. A thief may use a tool to open the locker to steal the valuables, or a master may use his dog to attack someone in this case the tool or the dog is not criminally liable, but the person who is using this instrument or animal for the crime is liable. In short, this model suggests that the programmer, user, owner, or supervisor will be criminally liable, whereas the AI has no criminal liability at all.

This theory is correct for a very basic level of AI where the AI does not have high cognitive power. Today's AI can decide to commit a crime based on its own accumulated knowledge, learning, and experience. Artificial General Intelligence (AGI) will be as intelligent as humans and Super AI will be even more intelligent than humans. So, in that scenario, it is not fair to hold the user or developer or the owner liable for the crime done by an AI.

3.2 Liability for Foreseeable Crimes Committed by Artificial Intelligence

In the second model, a little advanced version of AI is considered. For example, An AI developed for detecting viruses, malware, and spyware, itself becomes spyware and does spying and sends viruses to other systems by using its basic programming.

Here, the developer had developed the AI for a different purpose so the developer comes to know about the crime only after it is committed using the same program by an AI. In this model, the deep involvement of the programmer or the user is there but there is no intention to commit any crime by using AI. Similar to the case mentioned above of the Kawasaki factory, the AI robot considered the person who entered to repair it as a threat to its mission and killed him. The AI robot was not designed to kill humans, nonetheless, the human worker was killed as a result of the AI entity's action. AI changed the purpose of programming. In such cases, the first model cannot be applied as in the first model the user or the developer used an AI to commit the crime as an instrument with a criminal intention i.e. with mens rea. Whereas in the examples mentioned above in the second model, the developers or users did not have any intention to commit a crime, but the developer or user were negligent as they are required to know as a reasonable person that such an offence is a natural, probable consequence of their action. In such situations, the second model can be used. A person may be held accountable for an offence if that offence is a natural and probable consequence of that person's conduct. This is the fundamental principle of criminal law in case of negligence. The reasonable developer or the user of AI

should have known the probability of commissioning such an offence, he should have foreseen the offence and could have prevented it.

There can be two types of negligence. First, the developer or the user never had any criminal intention but they were just negligent. As in the example mentioned above the developer or the user should foresee an AI developed to detect spyware can become spyware itself and damage other systems. The programmer can be held liable for the cybercrime for negligence. The second situation can be when the developer of AI or the user developed or used AI intentionally to commit one type of crime but AI by mistake or by using the software differently committed another crime in addition to or instead of the planned crime. For example, an AI developed to steal things from houses, while committing the offence of theft if it kills who resists it from committing theft then it is the natural and probable consequence of the action and can be foreseen. For this, criminal negligence liability is not enough. It is far more than negligence alone. In such cases, they are to be held liable for the actual offence committed even though it was not planned by the user or the developer. So, in the above example, the developer or the user can be held liable for theft if committed as per the original plan as well as murder which was not planned.

3.3 The Direct Liability Model- AI As A Legal Person

To make any person criminally liable the presence of 'actus reus' and 'mens rea' is sine qua non. If an AI fulfils both these conditions then there is no reason for not holding an AI directly liable for the crime committed. If a robot AI assaults someone with the movement of its hydraulic arm it satisfies the condition of actus reus. Same way if any duty is assigned to an AI entity and if it does not perform it can be held liable for the actus reus of omission. The real challenge lies in holding AI mentally liable for a crime. To prove that AI had the mens rea i.e. knowledge or intent for committing a crime. Humans receive data from different sensory organs like eyes, ears, tongue, nose, and skin and then process the same in the brain and the result is its behavior or how a person acts. Advanced AI technology does the same work. They collect data from various sources. Learns from it, processes, analyses, and decides its future course of action. Even AI can think faster and better than humans. So, what is the logic for exempting AI from criminal liability? The AI and humans may be joint perpetrators and both can be punished accordingly

So, the criminal liability of an AI as per the third model of direct liability is the same as humans. The same criminal laws are to be applied to AI maybe with some minor modifications or adjustments.

3.4 Harmonization of Three Models

These three models are not mutually exclusive. It provides a guideline on when to use which model. When an AI is used just as an innocent agent or just as a tool and the real perpetrator is the developer, the user, or the owner, then the first model is to be applied. Under this model, it is assumed the agent does not have the capacity to do wrong and therefore the agent is to be considered innocent, and the person who acts through an agent is to be held liable. If a goods delivery agent is appointed to deliver goods from one place to the other. He may not be aware of what is inside the parcel. The parcel may contain illegal weapons or drugs. In such cases the agent is innocent but the person who made his agent deliver such goods will be criminally liable. Same way, the person who developed AI or used AI for criminal activity will be held criminally liable and AI will be considered an innocent agent who was just used as a tool to commit the crime. In the same case if the developer is not a human but another AI then that AI as a perpetrator can be punished as per the third model of direct liability of AI.

This model can be used only when the user or developer of AI is aware of the outcome and intentionally used AI to commit a crime but when the person using AI himself is not aware or does not have intention or mens rea for that matter the individual cannot be held criminally liable. If the possible outcome is reasonably foreseeable then at the most the user or the developer can be held liable for negligence if there is no mens rea for committing a crime under the second model of probable and natural consequence. Here also if the developer is also an AI then the third model of direct liability in addition to the second model may be applied. The harmonization of these three models can create an inescapable trap for the perpetrators whether the perpetrator is a human, a

corporation, or an AI. In turn, it will enhance the feeling of security in society and faith in the criminal justice system.

4. Punishments for AI

In the first instance, the idea of punishing an AI seems ridiculous, but actually, it is not. First, we need to understand what is a punishment.

In the words of H.L.A. Hart, punishment involves five elements: first, it involves pain or other consequences normally considered unpleasant. Secondly, it must be for an offence against the legal rule. Thirdly, it must be of an actual or supposed offender for his offence. Fourthly, it must be intentionally administered by humans other than the offender, and fifth, it must be imposed and administered by an authority constituted by a legal system against which the offence is committed

The debate is gaining momentum worldwide in favour of directly punishing AI for the crimes independently committed by it as well as where no human is directly involved. According to Gabriel Hallevy, "When an AI entity establishes all elements of a specific offense, there is no reason to prevent the imposition of criminal liability upon it for that offense." He vehemently advocates criminal liability for AI and is considered a flag bearer of this concept.

For a moment if we think, AI is tried, prosecuted, convicted, and held criminally liable. Now, the next question that arises is how to punish AI. How can it be given the punishment of imprisonment, death penalty, or fine? In most cases, AI does not have a physical body like an AI robot then whom to arrest and whom to imprison. The AI may not have monetary power or a bank account to pay a fine. Similar questions arose when the criminal liability for corporations was recognized. Some adjustments had to be made in punishing a corporation for crimes the same way some adjustments are needed for punishing an AI.

The major punishments under criminal law are capital punishment, imprisonment, community service, victim compensation, and fines. These punishments can also be inflicted on an AI with some minor adjustments without compromising the purpose of that punishment.

The object of capital punishment is to incapacitate the criminal and to stop the offender from committing any more crimes in the future, by depriving him of his life. Same way, various punishments can be given to AI to incapacitate or prevent AI from committing crimes in the future like deleting the software governing AI, or valuable and expensive AI can be forfeited, confiscated, or banned from the use or destruction of AI. Additionally, the AI owners can be heavily fined or imprisoned for the deterrence. The developers or owners may be asked to compensate the victims.

In most jurisdictions, imprisonment is considered a widely used form of punishment. Imprisonment deprives the criminal of his personal liberty of free movement. The personal liberty of an AI can be taken by suspending its use for a certain period of time.

In modern-day punishments, community service is gaining momentum. It means a compulsory contribution of labour to the community by the criminal. This punishment can be easily inflicted on AI. The AI which was used for profit generation may be used for community service.

Another form of punishment is fine. The purpose of a fine is to deprive a person of some of his property to create a deterrence effect. This is highly used in the case of criminal liability of a corporation as a punishment. Humans and corporations both can have assets and can have bank accounts. So, it is easier to deprive both of them of their properties. Whereas AI would not have property, bank account, or money. A small adjustment is to be made in inflicting fine on AI. Humans and corporations have acquired these properties through labour. When a fine has been imposed the ownership of the property which a natural person or a corporation acquired using labour is transferred to the state as a punishment. So, AI can be imposed to contribute labour in the form of fines for the betterment of society. Members of the European Parliament have called for "a mandatory insurance scheme and a supplementary fund" to ensure that the victims of driverless cars are sufficiently compensated

Similar insurance schemes can be designed to ensure that an AI offender has the means to pay its fines. It can be summed up that all major punishments of humans can be inflicted upon AI with minor changes.

5. Applying Theories of Punishment To AI

In criminal law, punishment is any pain, penalty, suffering, or confinement inflicted upon a person by the authority of the law and the judgment and sentence of a court, for some crime or offense committed by him, or for his omission of a duty enjoined by law.

According to Salmond, "Crime is an act deemed by law to be harmful to the society as a whole though its immediate victim may be an individual."

Crime is a serious evil and prevention of crime is the primary duty of the State. This can be achieved through inflicting punishments on the wrongdoers. Punishment can be used as a method of reducing the incidence of criminal behaviour either by punishing the offenders severely and setting an example for potential criminals to abstain from crimes, incapacitating and preventing them from doing further offenses, or by reforming them to be better individuals. We can apply theories of punishment to AI.

5.1 Deterrence Theory and AI

It is believed that if AI is punished it will not deter other AIs from committing crimes as AI is undeterrable. So, punishing AI does not serve the purpose of deterrence. Here, we must bifurcate specific and general deterrence. Specific deterrence is for the criminal who is punished. He would be deterred from committing crimes in the future. The undeterrability characteristic of AI may be true for basic AI as they are not designed to be sensitive to punishments or criminal laws, but future responsive AI collects data from various sources and learns from past experiences independently. Punishing AI will set an example and theoretically may have a deterrent effect. The general deterrence effect of punishment is that it prevents others from committing such crimes as this punishment sets an example for other potential criminals. It may or may not directly deter an AI but will create a general deterrence by discouraging other potential offenders from committing crimes using AI. It can be a deterrent not only for AI but also for other AI developers, users, etc. There can be punishments like heavy fines, and the destruction of AI (death penalty for AI) will definitely set an example to deter the developers, and inventors from developing such harmful, criminal AIs. As it requires huge cost, time, and effort to develop an AI. Destruction of AI would be financially disastrous for its developers, owners, users, etc which will encourage them to produce only socially beneficial AIs.

5.2 Retributive Theory and AI

Retributive means to 'payback'. The base of this theory is vengeance. By punishing the criminal, the victim gets a sense of satisfaction and does not take the law into his own hands to punish the criminal himself unlawfully. The victims of AI-related crimes will get a sense of justice when AI is punished which will increase the confidence of the society in the legal framework. The people will have the assurance that even the state has a zero-tolerance policy for crimes even if it is committed by an AI. This will create an overall environment of safety and security. On the contrary, if these robots or AI go unpunished it will create an atmosphere of fear that day by day AI is getting powerful, and in the near future, it will overpower humans.

5.3 Preventive Theory and AI

The purpose of this theory is to disable the criminal to prevent him from repeating the crime. Punishing AI can serve the purpose of the Preventive theory in the best possible manner by banning the use of this criminal AI or destroying it.

5.4 Reformative Theory and AI

The AI does not have a heart to be reformed. The anthropomorphism of AI for reformative theory seems irrelevant in the current scenario. Maybe future AI with emotions may learn from their punishment and may be reformed to not get involved in crimes, but that seems far from real today.

It is indeed a very challenging task to amend laws to hold AI criminally liable. An in-depth study of the pros and cons needs to be carried out as once these laws come into force they will have irreversible effects.

6. Mens Rea and AI

As per the general rules of criminal law, "99 Culprits Can Escape but One Innocent Shouldn't Be Punished". Same way the punishment should be in proportion to the crime. No one should be punished severely for a petty offense. The most important concept is "culpability" which means 'responsibility for a wrong'. That is why in any crime the courts do not see only 'actus rea' but the 'mens rea'- the guilty mind. As per Section 82 of the Indian Penal Code, 1860 anything done by a child under seven years of age will not be an offense as they do not understand the consequences, and cannot have the guilty mind or mens rea and so cannot be held culpable. A tsunami may create huge harm and destruction but it does not have a guilty mind. It cannot think. So, here the question is AI may not know the consequences of its actions. So how to consider an AI guilty?

AI does not have mens rea like intent, knowledge, or recklessness. So, to convict AI is against the set principles of criminal law requiring mens rea as the prerequisite for any crime. It is against the rule of law. Without mens rea the AI cannot be held culpable. There are various solutions for this.

General Artificial Intelligence (GAI) or Strong AI is capable of doing any intellectual job as effectively as a human being. Super AI is a level of system intelligence where AI is capable of outperforming humans in any task thanks to their cognitive abilities. Super AI is still only in theory but looking at the current development it will be a reality very shortly. Even GAI can collect data from various sources, analyze it, and can also find the best way of doing something independently. In this, it may not be in a position to understand many times which act is socially advantageous and which is against society. It may behave differently from the purpose for which it was created and it is difficult to find why it behaved in this way, and what was going on in its mind. This happens even with humans many times. Judges do not have clear ideas about the mental state of the defendant and infer from the behavior of the person about his mental state. In the same way, the judges can infer the intention of an AI from its behavior. If an autonomous car runs over someone, its behavior can be checked. If it changed its route repeatedly to target a particular person, that means the AI had an intention to kill that person. The intention of AI has to be checked from the programming level which requires expert opinion and may be quite costly.

7. Extension of Corporate Criminal Liability Concept to AI

Corporations are also artificial judicial persons and they are criminally liable. It took hundreds of years to bring corporations under the realm of criminal law for the wrongs done by them. A corporation does not have 'mens rea' but still they can be punished along with the directors if the activities are harmful. Recently, in Iridium India Telecom Ltd. v. Motorola Inc., the question came up with the Supreme Court of India regarding punishing corporations for crimes of cheating and criminal conspiracy. The Supreme Court, held that a corporate body can be prosecuted for cheating and conspiracy under the Indian Penal Code.

Also, in the cases of 'Strict Liability' a corporation can be held criminally liable. A corporation has tax liability apart from its promoters and directors and if any default is made the corporation itself can be held liable. In India even deities have legal personhood so why can't an AI have personhood?

Various countries do recognize corporations as a person and hold them criminally liable.

The corporation can be held criminally responsible for a variety of Crimes like Conspiracy, public nuisance, Violations of Consumer Protection laws, The Illegal practice of Medicine, Antitrust law Violations and many more.

So, with the changing times criminal liabilities for corporations or even for robots and AI is not an out of the box idea. An analogy between corporations and AI as legal entities can be easily drawn and criminal culpability for corporations can be extended to AI. The Republic of Korea has already started robot taxation by increasing the tax on the automation sector by reducing tax credits by 2 % points. In every country including India, corporations are being taxed and they are considered artificial judicial persons. The corporations cannot think

but still, they are considered a person for many reasons like taxation and are held guilty as well as punished whereas the AI can think and learn independently from the experience and various data sources so it must be given personhood and can be held criminally liable. The laws need to be amended for AI. The way Saudi Arabia gave citizenship to a robot named 'Sophia' developed by Hansson Robotics many more countries will be giving personhood and eventually citizenship to the AI

One of the ways to punish AI can be to punish its developers, owners, or users. As in the case of any crimes committed by the company the directors or Key managerial Personnel (KMP) may be held liable because a company being an artificial person cannot make decisions on its own.

The problem with this solution is that the company is not autonomous. It is completely dependent on its Board of Directors. The directors make decisions for the company. So, it is easier to hold the directors or the KMPs liable. Whereas in the case of AI most of the time the AI is autonomous and can make independent decisions and can also behave differently from the purpose for which it was developed. In such cases, it is unfair to hold the developer or the owner liable.

8. Strict Liability for AI

Another solution to hold AI culpable without mens rea can be similar to the cases of Strict Liability. In strict liability cases, culpability is not a necessary sine qua non. Strict liability crimes do not require a particular state of guilty mind. The best example of this is the Bhopal Gas Tragedy case. A new set of strict liability offences for AI crimes can be defined which an AI can commit without any mens rea. The concept of 'No fault liability' can be extended to AI without compromising the principle of legality of punishing AI sans mens rea.

Strict liability crimes are disfavoured by many legal scholars because it is unjust to punish someone who had acted without mens rea and many times in such cases the question of violation of human rights arises as a person is punished under strict liability without any intention of committing it. Such problems of violation of human rights do not arise in the case of AI.

There is a challenge in extending the concept of strict liability to AI. For a person to be guilty under strict liability, the person must have done that act voluntarily. For any criminal liability, a voluntary act is a must. Section 39 of IPC, 1860 states that "A person is said to cause an effect "voluntarily" when he causes it by means whereby he intended to cause it, or by means which, at the time of employing those means, he knew or had reason to believe to be likely to cause it". So, here the pertinent question is if AI cannot have mental states and cannot think about the consequences can the act of AI be considered voluntary for holding it criminally liable? To avoid this confusion some duties with high standards may be imposed on AI, omission of which can hold AI liable for crimes even without mens rea.

9. Arguments In Favour Of Punishing AI

Punishment is justified only if its benefits outweigh its cost.

9.1 Autonomous AI

There are cases where AI is making independent decisions, without human intervention or very trivial intervention of humans. The AI with its cognitive abilities may act in a completely different way from the purpose for which it was designed. In such situations, it is unjust, unfair, and unreasonable to hold the developer, owner, or user criminally liable. This is a strong case for punishing AI itself.

9.2 Enforcement problem

There may be an individual responsible for AI crime, but the individual is in the dark web and not identifiable. Some hackers have inserted the virus in the AI and have remained anonymous. Or the AI itself has created some virus.

9.3 Unfair to punish an individual where he has no role to play

To develop one AI there may be thousands of researchers, programmers, or developers are involved. They may have worked for several years in research and development. So, it is extremely difficult to reduce criminal liability to these people for the crimes committed by AI. It is extremely difficult to decide the contribution of each person. As per criminal law, the punishment should be in proportion of crime but it is impossible to decide the proportion or contribution of these people. They might not have created this AI for criminal activities and after years of proper behaviour, AI may change its behaviour through its learning from heterogeneous sources and gets involved in crimes. In such cases, no individual can be held liable for the crime committed by an AI but only and only AI should be punished.

9.4 Many AI crimes may go unpunished

This situation is very dangerous. As the current criminal Justice System does not recognize AI as a criminal and when no individual can be held liable for AI-generated crimes, these crimes will go unpunished as in the case of RDS mentioned above. This can create an environment of fear in the society.

9.5 Encourage research and development

If for autonomous actions of AI, the developers are punished then it will be a huge discouragement for new research and there will be a feeling of apprehension in new research and development of AI, which is not good for a country to grow. Instead, if the AI is directly punished for its autonomous actions then the researchers, inventors, and developers can develop better AI in the future without any apprehension.

9.6 Enhances faith in the criminal justice system

If AI-generated crimes are not punished then it will send a message that these kinds of crimes are permissible. If it is permissible for an AI it may become permissible for humans too. On the other hand, by punishing an AI the state can set an example that the state has a zero-tolerance policy towards crimes and criminals whether the criminal is a human, a corporation, or an AI. This will enhance the faith of the people of the nation in its criminal justice system which will help to maintain law and order. Also, labeling the AI as a criminal satisfies the vindictive desire of the victims and stops him from taking the law into his hands to take revenge.

10. Challenges

There are many benefits of punishing AI, even the basic theories of criminal laws can be extended to AI, but one must do a cost-benefit analysis before amending any law for AI. There are humongous challenges to hold AI criminally liable.

10.1 Significant amendments in laws

First and foremost, it will require hundreds of laws to be amended which is a mammoth task. In a country like India, and any other democratic country following a parliamentary form of government this will be a time-consuming process. It's a novel idea and so many stakeholders may protest and public awareness about the same has to be created before implementing it.

10.2 AI Legal Personhood

To hold AI liable for any crime first of all AI should be recognized as a person. The jurisprudential definition of a person is to be amended. Whether AI is to be given citizenship or not and if yes what kind of fundamental rights, statutory and legal rights will be available to it, and what will be the duties of AI?

In this direction, in 2017 the European Parliament asked the European Commission to create a legislative instrument to deal with the 'Civil Liability of Robots' It recommended giving electronic personhood to robots for attaching tort liability to them. This was just a tentative proposal but it was highly opposed by the stakeholders. More than 150 AI experts wrote an open letter to the European Commission not to give legal personhood to robots from an ethical and legal perspective.

10.3 Moral and Ethical issues

It will be very difficult to digest that machines are at par with humans and there are possibilities that they surpass human intelligence if the development of AI is not stopped at a reasonable level. The AI may overpower humans.

AI is capable of having the same type of antisocial behaviour as human criminals but without moral blameworthiness. AI just works on pre-installed programs. Even if AI goes out of control or behaves independently or performs different tasks than what was assigned to it, that is only because of faulty programming or lack of knowledge that what are the different ways in which AI can perform in the future. The people favoring punishing AI may still argue that it is important to punish the AI for its antisocial behaviour and blameworthiness plays a secondary role in deciding criminal liability for AI.

10.4 Balancing Rights and Responsibilities

It may be argued that rights and responsibilities or duties are two sides of a coin. If AI is considered a person and can be punished like a person then morally it should be given rights like a person. The AI is intelligent enough to vote. The AI may be given benefits under labour laws. Here, again it can be argued by drawing an analogy with the companies. The companies enjoy legal personhood but a company is not granted voting rights or fundamental rights or right to marry etc. The same strategy can be applied to AI too.

10.5 Financial and Research Aspect

As a severe punishment the use of that particular AI may be banned or that AI may be destroyed. In this case, the developer of that AI is punished indirectly who may be absolutely innocent. He might have done a lot of research and might have spent a considerable amount to develop this AI. And afterward, the AI may be behaving independently and might have committed a crime. This will highly demotivate other people to do research and develop new AI.

10.6 What if AI is a victim

If AI is considered a person and can be punished then can AI be a victim of crime? What if some rival destroys AI, does it amount to murder? Or if AI is damaged or harmed can AI or its developer or user get victim compensation?

10.7 AI cannot be punished in the real sense

According to Hart "The punishment must involve pain, suffering and unpleasant experience" Even if an AI is destroyed or reprogrammed it cannot feel bad or have unpleasant experiences or suffering. So, punishing AI does not serve the purpose in the real sense. However, it can be argued that not all criminals find their punishment unpleasant. Some may go to jail for publicity stunts or a poor person not having enough food may commit a petty crime to be in jail to get food and shelter.

10.8 For any act of AI, some human is ultimately responsible

A school of thought would argue that there is no need to punish AI as always there is some individual responsible for the crime committee by AI like its developer, user, owner, supervisor, etc. So, the law can easily punish these humans instead of making things complicated by punishing AI. The same arguments were given earlier for corporate criminal liability and also that some individuals are always culpable for the wrong done. For example, If a company's activities are creating pollution beyond a reasonable standard then instead of holding the company liable the Board of Directors or any other individual responsible for the same like safety manager or pollution control officer, etc can be punished. In the case of autonomous AI, the role of individuals may be missing at all or may be negligible. Anthropomorphism of criminal liability is a lacuna in the criminal justice system. It is not necessary in every case a human is a criminal especially when AI is involved. Whenever AI commits a crime, to punish a human for his minute misconduct or slight negligence is absolutely unfair, just to satisfy the requirement of the current legal system that only humans can be punished. In such situations instead of punishing a human, an AI must be punished directly. Sometimes nobody may be held liable for the

crime. In the United Kingdom, a ferry sank in the water and almost 200 people died. The official inquiry held that the company running a ferry must be liable but there were too many layers of responsibility and it was impossible to hold any one person liable ultimately the court did not hold any individual liable for this mishap.

10.9 Cascading Effect

When corporations are inflicted with a high amount of penalty for any crime, ultimately it is a punishment for innocent shareholders as this penalty reduces the divisible profit of the company. This is against the basic principle of criminal law that a punishment should be proportionate to the wrong done. When an AI is destroyed or reprogrammed, the AI does not suffer any pain but the cascading effect of this punishment is on its developer who might have spent a huge amount of resources and time in developing a unique AI. He may be completely innocent when AI commits a crime autonomously.

Against this, it can be argued that this cascading effect of suffering is not only in the case of AI or corporations. It is always there in any punishment even for humans. When a criminal is punished who is a bread earner of the family then due to his punishment the entire dependent family suffers. So, it requires reform in the criminal justice system and it is not unique to AI punishment.

11. Cost Benefit Analysis of Punishing AI and Remedies Other Than Punishment For AI- Generate Crimes

The punishment must be justified. Just because it prevents or deters or sets an example for others or gives the feeling of revenge it cannot be inflicted. The punishment is justified when there is no better substitute available than punishing, including doing nothing. There are many points in favour of punishing AI but at the same time, the challenges are huge. So, what can be the way forward for the time being? A proper cost-benefit analysis needs to be done to find out whether the benefits of punishing AI are more than the cost, time, and effort involved in setting the system for punishing an AI. There can be other remedies apart from punishing AI for crimes. An attempt is made here to check out such other remedies.

11.1 Expanding the scope of existing criminal law.

This is the simplest alternative remedy for punishing AI. In the current criminal justice system humans are held responsible for any crimes committed by any machine or computer. The machines are just considered a tool and not the perpetrator. For example, if a hacker tries to hack sensitive government data, then the human hacker will be held criminally liable for the crime and will be punished, not the software that was used to hack the data. For punishing humans every country has enough penal laws and cyber security laws under which these humans using machines or computer programs can be held responsible for the crime. For AI- generated crimes the scope of these existing laws can be expanded to define new crimes specific to AI. This situation can be compared with the ‘Innocent Agent ‘model.

The complex situation arises when the person has used AI to commit a minor crime, for example, to steal data from the computer of a police station but the AI erroneously or autonomously damages the property and causes the death of certain people whom it considered can stop it from accomplishing his work. Can the hacker be held liable for the murder of these people which was not foreseen by the hacker or he did not have a mens rea for murder? For such cases, criminal laws already have many provisions like constructive liability concepts. These concepts just need to be expanded to accommodate AI-generated crimes in its realm.

In the current situation when AI-generated crimes are very less the best solution can be to define new crimes related to AI, the way cyber crimes were defined when computer-related crimes increased. A new AI Crimes Act can be enacted to criminalize negligent or mala fide intentional use of AI by different stakeholders. The individuals related to AI at different stages like developers, users, owners, supervisors, and trainers can be punished for the irresponsible behaviour of an AI.

Again, this remedy is not foolproof, This can work only in the current situation when AI is not fully autonomous. In November 2023 it is believed that Open AI CEO Sam Altman had already developed GAI named Q* which is as intelligent as humans. If hypothetically, such an AI which is developed for socially

beneficial purposes may get information on its own from the dark web and enter into socially harmful activities then AI is no more an 'innocent agent' but the developers are the innocent agents. For such AI the developers or users cannot be punished under the realm of existing criminal law and AI cannot be considered only a tool, else it will be a huge demotivation for the developers to invent new advanced AI.

11.2 Mandatory Licensing and Registration

There may be a requirement to register a designated responsible person before using AI who can be held criminally liable for any wrongful conduct of an AI. This person can be a natural person or an artificial person like a company or an NGO. The registration or getting a license before developing or using AI should be made mandatory. This can again be a difficult task. The license issuing authority must have AI experts as officers who can understand the criminal capability of an AI before issuing the license. It is very difficult to hire such highly technical staff, especially in developing nations. The cost of training and establishing the mechanism for issuing such licenses will be very costly compared to the benefits it can achieve. So, ultimately this remedy also infers punishing a person associated with AI and not AI directly.

11.3 Civil Liability

For AI-generated crimes, the scope of civil liability can be expanded, especially under the law of tort the concept of negligence and product liability should be elaborated. Negligence normally requires a lack of reasonable care or reckless behaviour, without that no damages can be recovered. This requirement can be removed in the case of AI. In the same way, product liability requires both an AI to be a commercial product and the product must be defective. This should also be amended for AI. As AI is not always a commercial product it can be a service and in the case of advanced AI it will be very difficult to find defects in AI. AI can commit a crime without any defect also. There can be insurance against AI-generated crimes. The registered person can save himself from civil liability by taking insurance.

11.4 Moral Algorithms and separate Criminal Code for AI

The developer has to encode all moral criteria in advance. Set such standards that the AI learns moral values from these standards. Some examples can be fed into its system in the ethical uncertain situation, what can be the best moral behaviour. For example, 'The Medical Ethics Expert' (MedEthEx) is an ethical advisor that will guide healthcare workers in the ethical dilemma. Machine learning techniques are used to extract decision principles from cases with conflicting prima facie duties and apply these principles to determine the correct course of action in similar and new cases. An AI can be taught with reinforcement learning that if the accident is inevitable human life is more important to be saved rather than loss of property.

Some experts even want a separate criminal code for AI. As there can be higher moral standards for AI than for humans. If a person is drowning and a bystander does not save him he is not criminally liable. But if that bystander is a robot and there is a high probability that the robot could have saved the person the robot can be held liable. So, it is very important to define the scope of moral standards for AI. What is prohibited, permitted, or obligated for an AI morally, should be fed into the AI system. Therefore, a separate Criminal Code can provide the bare minimum moral standards for all AI to adhere to. For example, if a tiger attacks a person to satisfy its desire to have meat, we do not consider the tiger morally wrong but if a man or an AI robot attacks another man to satisfy his desire to eat meat will be morally unacceptable. There should be proper standards of moral responsibility for AI. Another example can be if a self-driven car loses control and an accident is inevitable. If it takes a left turn it will crash a poor child and if it goes right it will crash a wealthy woman. The AI makes an immediate calculation that crashing a poor child will lead to payment of less damages for its owner than crashing a wealthy woman and so it crashes the poor small child. The moral standard should be set for such a situation which is not based on the wealth of the victim. To avoid ambiguity such standards should not be decided by the AI manufacturers, developers, or users but by the society collectively. The people at large should decide what is morally allowed for an AI. It will also be a warning to AI developers to take proper care of socially accepted moral standards while designing an AI. This will help to hold these developers and manufacturers as well as users criminally liable if they do not abide by the set rules.

12. Conclusion

The way artificial persons like corporations are recognized as a person in the eyes of law and bear criminal liability, AI should also be given personhood and it should be brought into the ambit of penal laws. As AI is becoming more and more complex, advanced, and autonomous day by day it is going to be impractical to hold any individual criminally liable for AI-generated crimes. So, to punish AI is the best solution. But for the time being when AI is still in its nascent stage, General AI and super AI do not exist and not many AIs are directly involved in crimes, still, AI is more or less under the control of humans. At this stage, it is better to go for substitute remedies rather than punishing AI. Instead of directly punishing AI, the scope of criminal and civil penalties can be extended. There may be severe penalties on the developers, owners, users, or supervisors of AI if they are not responsible enough to manage their duties properly. This seems to be a better and easier solution than punishing AI directly and the stakeholders indirectly. Severe civil liability in the form of tort for failing to responsibly supervise, operate, design, or develop an AI can be introduced. As far as possible criminal liability should be less as it would discourage the invention and development of new AI which can be disadvantageous for society and the economy at large. This can be a feasible solution for AI-related crimes instead of punishing AI directly, not because it is against criminal law theories but because the cost will be very high compared to the benefits it can fetch in the present scenario. The same benefits can be achieved by increasing the civil and criminal liabilities of stakeholders, amending current laws, strict licensing and registration policies for AI, etc can be used to deter AI from committing crimes, or at least it will provide a deterrence for the owner, developer, and users of AI.

At the same time, it is not premature to consider the criminal liability of AI and punish it as scientists are continuously in the process of creating advanced AI that can make moral decisions independently. There is an urgent need to regulate and have binding international laws in the areas of self-driven cars, autonomous killer weapons, darknet, etc It is always good to be future-ready, otherwise in no time, AI can overpower humans.

References

[1] Greg Bensinger & Tim Higgins, Uber Suspends Driverless-Car Program After Pedestrian Is Killed, WALL ST. J. (Mar. 20, 2018), https://www.wsj.com/arti- cles/uber-suspends-driverless-car-program-after-pedestrian-is-struck-and-killed- 1521480386 retrieved on 3rd September, 2023.

[2] https://www.theguardian.com/theguardian/2014/dec/09/robot-kills-factory-worker,retrieved on 12th December,2023.

[3] The Reasonable Robot, artificial Intelligence and the Law, Ryan Abbott, Cambridge University Press, at 15

[4] https://ideaexchange.uakron.edu/cgi/viewcontent.cgi?article=1037&context=akronintellectualproperty, retrieved on 13th December, 2023.

[5] Supra at 2.

[6] Herbert Lionel Adolphus Hart, Punishment and Responsibility4-5 (Oxford Univ. Press, 1968). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3227710, retrieved on 14th December, 2023.

[7] Gabriel Hallevy, The Criminal Liability of Artificial Intelligence Entities: The Criminal Liability of Artificial Intelligence Entities- From Science Fiction to Legal Social Control, 4 Akron Intellectual Property Journal, 171,191(2010),https://ideaexchange.uakron.edu/cgi/viewcontent.cgi?article=1037&context=akronintellectualproperty, retrieved on 3rd December, 2023.

[8] European Parliament Press Release 20170210IPR61808, Robots and Artificial Intelligence : MEPs call for EU-wide Liability Rules (June 10, 2017, 13:09), https://www.europarl.europa.eu/news/en/press-room/20170210IPR61808/robots-and-artificial-intelligence-meps-call-for-eu-wide-liability-rules, retrieved on 14th December, 2023.

[9] https://ccsuniversity.ac.in/bridge-library/pdf/LLM-II-SEM-JURISPRUDENCE-II-L-2002-Lecture-on-Theories-of-punishment.pdf retrieved on 3rd December, 2023.

[10] https://main.sci.gov.in/jonew/judis/37034.pdf, retrieved on 16ht December,2023.

[11] https://www.warc.com/newsandopinion/news/south-koreas-robot-tax/en-gb/39118, retrieved on 16th December, 2023.
[12] https://www.wired.co.uk/article/sophia-robot-citizen-womens-rights-detriot-become-human-hanson-robotics, retrieved on 16th December,2023.
[13] https://www.scconline.com/blog/post/tag/bhopal-gas-tragedy/ retrieved on 15th December, 2023.
[14] https://www.europarl.europa.eu/RegData/etudes/ATAG/2017/599250/EPRS_ATA(2017)599250_EN.pdf, retrieved on 4th December, 2023.
[15] Open Letter to European Commission artificial Intelligence and Robotics, https://robotics-openletter.eu/ retrieved on 4th December, 2023, at 01:53 a.m.
[16] Punishment: Definition and Justification, Thomas McPherson, Oxford University Press, Vol. 28, No. 1 (Oct., 1967), pp. 21-27.
[17] Christian List & Philip Pettit, Group agency 167 (Oxford Univ. Press 2011), taken from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3227710, retrieved on 15th November, 2023.
[18] https://www.wired.com/story/fast-forward-clues-hint-openai-shadowy-q-project/ retrieved on 4th December,2023.
[19] https://www.researchgate.net/publication/229043957_MedEthEx_Toward_a_medical_ethics_advisor retrieved on 14th December, 2023.
[20] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3227710, supra at 522.