ON ROBOT CRIMES AND PUNISHMENTS

Priyam Jhudele[*]

"With artificial intelligence we are summoning the demon. It is the biggest existential threat to humans"

– Elon Musk

"The development of full artificial intelligence could spell the end of the human race."

– Stephen Hawking

Abstract

Humans have taken giant strides in the field of robotics in recent years. From self-driving cars to interactive software like Siri, terrific progress has been made. However, as we move forward towards smarter machines, machines that can think and react, we need to be wary of the threat they pose to mankind. This is the reason why renowned experts like Stephen Hawking and Elon Musk have expressed their concern over the increasing use of artificial intelligence. Moreover, the cases of robots committing crimes are on the rise, whether those crimes are committed directly or indirectly. We cannot impose liability for a crime committed by an intelligent thinking machine on a human

[*]Priyam Jhudele is a second-year student at the National Law Institute University, Bhopal. The author may be reached at priyam.jhudele@gmail.com.

merely because he programmed it. We need to devise a system of liability for punishing the intelligent machines which can think on their own. Thus, this paper proposes three models of liability under which artificially intelligent entities can be punished. It also proposes tests which can be used for determining their intelligence and the appropriate punishments which can be given to such entities. It also suggests changes which can be brought about in the Indian law to bring such entities under the purview of law.

***Keywords**: Artificial Intelligence (AI), Robots, Turing Test, Sentience, Accomplice liability*

I. INTRODUCTION

Humans have always been fascinated with creating things and they have always wanted to replicate their creation. Each of man's creations has been used in ways that can benefit or harm our species. So is the creation of robots. Robots that are not just machines but actual thinking entities. Machines that have artificial intelligence. This is the same artificial intelligence that we have witnessed in movies like *Her* or *The Matrix*. Watching the things that Artificial Intelligence ("AI") can do seems exciting in movies, but are we completely safe from these thinking machines? The worry expressed by SpaceX founder Elon Musk and renowned scientist Stephen Hawking must not be taken lightly.

What is more worrying is that incidents are actually taking place where AI entities are committing crimes, and in most cases, even getting acquitted because they are not subject to law. Recently, a

Volkswagen employee was crushed to death by a robot working in the same factory.[1] Now, interesting as it may sound, the persons who may face liability for the death may be the plant owner, the robot designer-hardware or software designer or *the robot itself*. A similar incident occurred in India where a robotic arm of a machine pierced a man who was adjusting the sheets on the machine.[2] Another surprising incident happened in Switzerland. A Swiss art group created a bot that performed automated shopping by purchasing several random items on the darknet.[3] What is notable, however, is the list of items that the bot purchased, which included pills of the drug 'ecstasy' and a Hungarian passport. The Swiss authorities arrested it, only to release it later.[4] Yet another incident involved a Dutch man whose Twitter bot tweeted a death threat to another bot. It tweeted "I seriously want to kill people at a fashion event". This led the police to question the man.[5] Moreover, in the United States alone, one person is killed by an industrial robot every year.[6]

Society devised criminal law in order to curb crimes by putting people in fear of punishment. It was the most effective tool for social control. But what about entities that are indifferent to the existing

[1]E. Dockterman, *Robot Kills Man at Volkswagen Plant*, TIME (Jul. 1, 2015), http://time.com/3944181/robot-kills-man-volkswagen-plant.

[2]R.J. Singh & S. Yadav, *Terminator redux? Robot Kills a Man at Haryana's Manesar Factory*, THE TIMES OF INDIA (Aug. 13, 2016), http://timesofindia.indiatimes.com/india/Terminator-redux-Robot-kills-a-man-at-Haryanas-Manesar-factory/articleshow/48460738.cms [hereinafter R.J. Singh & S. Yadav].

[3]A dark net is a private network with a restricted access which is not discoverable by usual means such as search engines.

[4]A. Toor, *Who's Responsible When a Bot Buys Your Drugs?* THE VERGE (Jan. 15, 2015), http://www.theverge.com/2015/1/15/7551031/automated-bot-buys-ecstasy-darknet-art-exhibit.

[5]M. Singleton, *Man questioned by police after his Twitter bot makes death threats*, THE VERGE (Feb. 12, 2015), http://www.theverge.com/2015/2/12/8025475/twitter-bot-police-death-threats.

[6]Accident Search Results Page, OSHA.GOV, https://www.osha.gov/pls/imis/AccidentSearch.search?acc_keyword=%22Robot%22&keyword_list=on.

legal system. Even the punishments we have devised for curbing crimes may fail. For example: will a robot feel the isolation that a human feels when incarcerated? Definitely not. Though we have developed effective thinking machines, the fact remains that they are not human and therefore punishments have to be devised such that they suit them. Gabriel Hallevy, a professor of criminal law and a frequently cited author in the Israeli Supreme Court, has tried to put forth a system of punishment on the same line as that of punishing corporate entities. His argument is based on the premise that, like robots, corporate entities were feared earlier- who was to be held accountable in case they committed any wrong and how were they to be punished? He thus believes that it should be robots that should be punished for their autonomous acts and not their programmers or designers.[7]

No discussion on the liability of robots goes without discussing Isaac Asimov's three laws of robotics which he introduced in his book *I, Robot*.[8] The laws read as:

> *1. First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.*
>
> *2. Second Law: A robot must obey orders given it by human beings, except when such orders conflict with the First Law.*
>
> *3. Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.*

A zeroeth law was later introduced in his book Foundation and Earth which read as:

[7]Gabriel Hallevy, *The Criminal Liability of Artificial Intelligence Entities—From Science Fiction to Legal Social Control*, 4 THE AKRON INTELL. PROP. J. 171 (2010) [hereinafter Hallevy].
[8]ISSAC ASIMOV, I, ROBOT (Spectra 1950) (2004).

> *0. A robot may not injure humanity, or, by inaction, allow humanity to come to harm.*[9]

These laws are not codified laws but merely guidelines on robot ethics and are a work of fiction. What these laws do well is that they prohibit robots albeit theoretically from harming human beings.[10] However, these laws gained tremendous popularity and have been used in a lot of robots since their inception.[11]

It is also imperative here to discuss the laws given by robotics physicist Mark Tilden, a pioneer in the field of robotics. His laws go as follows:

> *1. A robot must protect its existence at all costs.*
>
> *2. A robot must obtain and maintain access to its own power source.*
>
> *3. A robot must continually search for better power sources.*

If we go strictly by the word, Tilden's laws give a clear indication of what the robot species of the future could turn out to be. The first law indicates that a robot must protect its existence at "all costs" which means that a robot should protect its existence regardless of the damage it could cause to other living species or objects nearby. This is exactly what was exhibited in 1981 when a robot mistakenly identified an employee of a motorcycle factory as a threat to its mission and killed it with its powerful hydraulic arm and then resumed its duties with nobody to interfere in its mission.[12]

---

[9]ISSAC ASIMOV, FOUNDATION AND EARTH (Doubleday 1986) (2004).
[10]Sam Lehman-Wilzig, *Frankenstein Unbound: Towards A Legal Definition of Artificial Intelligence*, 13(6) FUTURES 442, 457 (1981).
[11]R. Murphy & D.D. Woods, *Beyond Asimov: The Three Laws of Responsible Robotics*, 24(4) INTELLIGENT SYSTEMS 14, 20 (2009).
[12]Y.H. Weng, C.H. Chen & C.T. Sun, *Toward The Human–Robot Co-Existence Society: On Safety Intelligence For Next Generation Robots*, 1(4) INT'L J. OF SOCIAL ROBOTICS 267, 282 (2009).

The question of how to control these intelligent machines still remains. Should we subject them to criminal law like we do for humans? If yes, then how do we effectively devise an inclusive code punishing such entities? Also, how do we plan to separate the actually intelligent entities from the rest? Further, even if we come to the conclusion that these AI entities are responsible for their acts, how do we punish them and will the traditional punishments be effective against them? Moreover, how do we plan to incorporate AI liability under the Indian Penal Code?

For this purpose, this paper proposes three approaches to AI liability. The three approach model was first proposed by Gabriel Hallevy in his work "I, Robot – I, Criminal".[13] This is the most favourable model for AI liability as it is inclusive of all possible ways in which we can hold AI entities liable. Moreover, the three approaches to AI liability should not be seen separately as all three approaches are essential for imposing liability on AI entities. The three approaches have been discussed in detail in Part II of this paper. Further, the tests to separate the intelligent machines from the non-intelligent machines have been mentioned in Part III of the paper. Part IV discusses various punishment adjustments that can be reached in order to punish such entities. The scenario under the Indian legal system is discussed under Part V of this paper.

II. APPROACHES TO AI LIABILITY

The primary focus of criminal law is on attributing liability to a person and punishing him for his acts. It has the dual requirement of *actus reus* and *mens rea*.[14] It is easy to establish actus reus, which is

[13]G. Hallevy, *"I, Robot – I, Criminal"—When Science Fiction Becomes Reality: Legal Liability of AI Robots committing Criminal Offenses*, 22 SYRACAUSE SCI. & TECH. L. REP. 1, 9 (2010).
[14]RATANLAL & DHIRAJLAL'S THE INDIAN PENAL CODE 3 (31st ed. 2007).

expressed in the form of acts and omissions. Mens rea, however, has various elements, such as knowledge, intention, and motive. A person can be subjected to punishment under criminal law only when the two elements exist together. Thus, there can be actus reus on part of a person but if there is no mens rea, he cannot be punished. Although, there are exceptions in the form of strict liability and rash or negligent acts.

In case of artificially intelligent entities, actus reus can be established in the same manner as for humans. However, the difficulty lies in assigning intention to these entities. The question has been getting difficult with the advancement of modern science and the development in thinking capabilities of these entities. Can an entity which can defeat Garry Kasparov in a game of chess not think of committing a crime?

To answer the same, this paper has adopted three approaches to liability of AI entities,

- Product Liability
- Consequence of an act
- Direct Liability

The classification would be as follows:

A. Product Liability

The first approach treats robot entities as merely products and thus encourages imposing liability on the manufacturers of such entities as is the case of all products currently in the market. This may fit well into the existing legal system. The liability of an act which the entity commits may be compared to that of a malfunctioning product with the onus on the manufacturer. Also, the other agents involved in importing, exporting and transporting such entities may be held accountable as and when necessary. Science-fiction writer Cory

Doctorow supports this view and writes "*For the life of me, I can't think of a law for robots that wouldn't apply to computers and vice versa."*[15] For law professor Neil Richards and Robotics professor Bill Smart, *"robots are, and for many years, will remain tools. They are sophisticated tools that use complex software, and no different in essence to a hammer."*[16]

Part of this approach also considers them as agents of the masters. While considering this concept, the common agent-principle relationship may be put to use. So, under this approach, it will be immaterial whether or not the agent is legally competent to commit the act. Here, the only thing important is that the relationship between the two forms is that of agent-principal. Even persons who are not *sui juris* can be agents. Furthermore, the law of agency require any formal acceptance of duty on behalf of the agent.[17] The rulings have even used 'automation' and 'human machine' to describe agents.[18] Meechem notes, anyone can be an agent who can perform the functions involved.[19] Even in common law a master is held liable if the act done by the servant was done to further the master's interest.

The Engineering and Physical Sciences Research Council (EPSRC) and Arts and Humanities Research Council (AHRC) published a set of principles which are to act as guidelines for all developers and users of robots. It would not be wrong to say that their principles are more or less inclined towards this approach of imposing minimum liability on the robot. One of these principles makes it clear that robots should be treated as products.[20] Being products, they must be

[15]C. Doctorow, *Why It Is Not Possible To Regulate Robots*, THE GUARDIAN (Apr. 2, 2014), https://www.theguardian.com/technology/blog/2014/apr/02/why-it-is-not-possible-to-regulate-robots.
[16]*Id*.
[17]W.A. SEAVEY, HANDBOOK OF THE LAW OF AGENCY 32 (West Publishing Company, 1964).
[18]*Id*.
[19]F.R. MEECHEM, OUTLINES OF LAW OF AGENCY 8 (Callaghan, 1903).
[20]*Principles of Robotics*, EPSRC.AC.UK,

subject to certain standards and safety measures. Thus, any hazard which is caused after that should be covered by the principle of caveat emptor, as people are well aware that a software may malfunction if hacked into. Another principle goes on to say that it is humans, not robots that are responsible for breaches of law. The breach of any law by a robot is the moral and legal responsibility of humans and not robots. It states, robots are just tools, designed to achieve goals and desires that humans specify. Users and owners have responsibilities as well as designers and manufacturers.[21] Absolving robots of all liabilities, the last of these principles go on to state that it should be made mandatory to register all robots and their manufacturers, user names and the liability on these humans rather than robots. From this approach, joint liability can be imposed on the user and the manufacturer where required.

However, there are several problems to such an approach. First, the peculiarity of such entities. Modern AI entities are not brought into existence by a single team. There are separate software and hardware teams. So, first the malfunction has to be traced back to the true manufacturer. Another problem with such an approach is that it is not forward looking, it treats AI entities as products which means that they cannot act on their own which itself is a wrong assumption. Modern AI entities are capable of performing several complex tasks of their own including programming themselves. The principle of inherent risk creates a problem here. A manufacturer can only be held liable if he fails to attach a warning to an inherently dangerous product. In case of a versatile device such as an AI entity, it is almost impossible to attach a warning to every task it performs.

B. The Natural-Probable Consequence

https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/.

[21] *Id.*

Consider a case where a robot is asked to perform a particular task and it is working diligently towards it; say a self-driving car is asked to reach a particular destination within a specified time. To reach on time, the car not only hits two people on the way but also goes well beyond the speed limit.

Here, the AI entity in form of the car never intended to hit anybody or intentionally drive beyond the speed limit. It was merely acting according to the instruction given to it by the owner. But neither did the owner want the car to hit anyone, he simply wanted it to reach there on time. Who is liable in such a situation?

What is common in these cases is that these robots were ordered to perform some specific tasks and they took extreme steps while performing the tasks conferred on them. Here, neither the robot intended to breach the law nor the owner who assigned it the task. Who is to be held responsible in such a scenario?

The Volkswagen case in which a machine grabbed a man and crushed him against a metal plate falls squarely under this category. The Swiss bot which ordered ecstasy pills would also fall under this. Interestingly, in the case of Swiss bot, the bot and the owner went scot-free as they could not be held culpable under the existing legal system.[22]

The situations explained above are not covered by the first approach simply because the first approach assumes intention on part of the programmers and the users of the bot whereas in the instant case, there is no prior intention on part of the programmer or the user to use the bot to commit the crime.

[22]J. Kasperkevic, *Swiss Police Release Robot That Bought Ecstasy Online*, THE GUARDIAN (Apr. 22, 2015), https://www.theguardian.com/world/2015/apr/22/swiss-police-release-robot-random-darknet-shopper-ecstasy-deep-web.

The natural-probable consequence doctrine was used to punish accomplice liability where the person who aids or facilitates the act of crime is as guilty as the person actually committing it.[23] The user of the AI entity can be punished for his negligence, even when the specific offense committed requires a different state of mind.[24] The American Law Institute's Model Penal Code also accrues knowledge where there is high probability of the offence being committed unless there is a clear conviction of the absence of any mishap.[25] Thus, reasonable programmers should be wary of the acts that the bots they create can commit. Hallevy[26] suggests that negligence in such cases can be attributed in two different conditions. One, where there is absence of any intent on part of the programmer to make the robot commit a crime and the crime is the outcome of a mere negligence while programming. Thus, in such cases, he argues that the programmer should be prosecuted under the law of negligence for the specific offence. While creating artificially intelligent entities, it should be prime concern of the programmer that a life-taking machine is not created. Next, he suggests, is when there is purposeful planning on part of the programmer to create an entity which commits a particular offence but the entity goes on to commit another offence. An example of this is the self-driving car knocking persons on the way discussed above. In such cases, the programmer shall be wholly accountable for the end result and knowledge presumed on his part.

The liability of the AI entity in such situations can be very well explained by Joshua Dressler's propositions on accomplice liability.[27]

[23]Michael G. Heyman, *The Natural and Probable Consequences Doctrine: A Case Study in Failed Law Reform*, 15 BERKELEY J. CRIM. L. 388 (2010).
[24]THE AMERICAN LAW INSTITUTE, MODEL PENAL CODE 312 (1962, 1985); State v. Linscott, 520 A.2d 1067, 1070 (1987).
[25]*Id*.
[26]Hallevy, *supra* note 7.
[27]Joshua Dressler, *Reassessing the Theoretical underpinnings of Accomplice Liability: New Solutions to an Old Problem*, 37 HASTINGS L.J. (1985) [hereinafter Dressler].

Dressler, a distinguished professor at the Ohio State University, focuses on the resultant social harm from the accomplice's conduct and suggests three theories for accomplice liability which are appropriate in our case.

First, Dressler suggested that accomplice liability might be based on one's substantial participation in the venture. If an accomplice spends substantial time on the criminal venture and his activity is significant in the completion of the criminal act, then the accomplice might be held liable. In our scenario, the robot entities, if they perform substantial part in the commission of the crime and are thus significant for the commission of the crime, they can be held accountable for accomplice liability. How much of what part of an act might constitute significant contribution is however, subject to the peculiarities of each case.

Next, Dressler examined the "control or hegemony test".[28] Under this approach only those who exercise control over the act are to be held accountable, however this approach may allow the AI entity to go unscathed and complete liability shall fall on the owner or the programmer. An example of this could be when a programmer programs an AI entity to commit a bank robbery, the entity is not programmed to break the electronic locks yet it improvises and goes on to commit the robbery. In such situations, this test roots for a lesser punishment for the entity because the control was exercised only by the programmer or the human agent involved. However, the improvisation was substantial in the successful completion of the robbery. This approach is thus not preferable.

Finally, Dressler considers a causation test, where he divides accomplices into "causal accomplices" and "non-causal accomplices" the terms clearly explaining what he means by them. This approach is also not preferable. Thus, the first approach considering substantiality

[28] *Id.* at 124.

of the accomplice's participation in the offence is most preferred and should be used in order to determine AI liability under the second liability model.

C. Direct Liability

In the case of direct liability, the AI entity is held directly liable. What makes punishing such entities difficult is that the dual requirement of mens rea and actus reus needs to be met. In case of AI entities, the actus reus or the physical part of the act can be easily established from the situation. However, what is difficult is establishing mens rea on part of such entities. In the case of strictly liable offences, punishing them is even easier as there is no need to establish mens rea.

Mens Rea is essentially the mental part of a crime which an individual needs to possess to be convicted for the crime. The Latin phrase "*actus non facit reum, nisi mens sit rea*" which means that act in itself is not a crime unless done with a guilty mind explains the importance of establishing mens rea. There are several levels of mens rea identified by different jurisdictions. The primary among them being, intention, knowledge and negligence.

Intention is the conscious exercise of the mental faculties of a person to do an act, for the purpose of accomplishing or satisfying a purpose.[29] All such provisions which require 'voluntarily', 'wilfully' and 'deliberately' doing an act are fulfilled with the fulfilment of the requirement of knowledge. An AI entity has mental faculties available to it and can consciously exercise it, as we have seen in the case of the 1981 killing by a Japanese robot which purposefully killed a man who was a 'threat' to its mission.[30] A definition more suitable to our cause has been laid done by the Indian judiciary in the case of *S. Raghubir*

[29]PSA PILLAI'S CRIMINAL LAW 7 (Lexis Nexis, 2008).
[30]*Trust Me, I'm a Robot*, THE ECONOMIST (June 8, 2006), http://www.economist.com/node/7001829.

Singh v. Commissioner of Income Tax[31] wherein the court said that intention is the fixed direction of mind to a particular object, or determination to act in a particular manner. If we are to treat modern day thinking machines as subject to law, their cognitive capabilities are not any less than compared to thinking humans. An example of this is the recent case of Google's Deep mind defeating professional Go player Lee Se-Dol.[32]

Knowledge is awareness on part of the person concerned, indicating his mind.[33] A person can be supposed to know when there is a direct appeal to his senses.[34] AI entities are well equipped to receive information are capable of perceiving and deducing aspects of the world. They have ability to deduce by using sensors such as cameras, microphones, tactile sensors, sonar and other sensors.[35] This concept is called in technological parlance as machine perception.[36] Modern machines can even simulate how the human mind works and synthesise its responses.[37] This is quite similar to what humans do, they receive data through their sensory organs, eyes, ears etc. and then process it to respond accordingly.[38] Thus, for offences which require knowledge, AI entities can easily be convicted therein.

[31]S. Raghubir Singh v. Commissioner of Income Tax, A.I.R. 1958 Punj. 250 (India).
[32]S. Byford, *Google's DeepMind Defeats Legendary Go Player Lee Se-dol*, THE VERGE (Sept. 3, 2016), http://www.theverge.com/2016/3/9/11184362/google-alphago-go-deepmind-result.
[33]Dressler, *supra* note 27.
[34]H.S. GOUR, THE PENAL LAW OF INDIA 240 (11th ed. 1998).
[35]STUART J. RUSSELL, PETER NORVIG, J.F. CANNY, J.M. MALIK & D.D. EDWARDS, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH (2d ed. 2003).
[36]NILS J. NILSSON, ARTIFICIAL INTELLIGENCE: A NEW SYNTHESIS (Morgan Kaufmann Publishers, 1998).
[37]PAMELA MCCORDUCK, MACHINES WHO THINK (2d ed. 2004).
[38]Daniel C. Dennett, *Evolution, Error, and Intentionality*, THE FOUNDATIONS OF ARTIFICIAL INTELLIGENCE 190 (Derek Partridge & Yorick Wilks eds. 2006); B. Chandrasekaran, *What Kind of Information Processing is Intelligence?* THE FOUNDATIONS OF ARTIFICIAL INTELLIGENCE 14 (Derek Partridge & Yorick Wilks eds. 2006).

If the elements of knowledge and intention can be fulfilled by AI entities, negligence can certainly be fulfilled as well. Negligence is the gross and culpable neglect or failure to exercise that reasonable and proper care and precaution to guard against any injury either to the public generally or to an individual in particular.[39] AI entities which are programmed to perform tasks efficiently should not bypass the laws and do harm to others through their negligence. However, in the cases of negligence, the liability would fall more on the programmer than the entity itself.

Many scholars have argued that although it may be possible for the AI entity to replicate the human mind, but it is not possible for it to display emotions like humans do. However, this will not be relevant in punishing AI entities because there is no requirement of an emotion to punish a person. If both the mental and the physical elements are present, it is only appropriate to punish such entities. Although it may not have an effect on the offence part as the requirements of offences are fulfilled, it may certainly have an impact on defences. It would mean that there will not be any defence available to such entities. Nobody expects a machine which can defeat the best chess and 'Go' players in the world to commit mistakes. Thus, the traditional defences like intoxication and insanity do not apply to AI entities. What is needed is an adjustment or introduction of new defences for this purpose. The AI entities can likewise be given a defence for any intrusion or corruption in their software. A virus attack on their system might have a corrupting effect and thus should be allowed as a defence.

Thus, AI entities should be made subject to punishments and liability imposed on them. It is imperative to note here, however, that the liability of the AI entity itself does not mean that the programmer or user goes scot-free.

[39]Empress of India v. Idu Beg, (1881) 3 I.L.R. 776 (All. India).

D. Interplay Between the Three Models of Liability

The abovementioned models of liability are not mutually exclusive and are to be read together to form a scheme of liability where punishment can be doled out according to the level of participation in the commission of crime. Moreover, it should not be that in imposing liability upon AI entities, people start using it as a shield to commit crimes and then escape punishment, the whole burden falling upon the AI entity. The cases where each of these models can be used has already been discussed above. The self-driving car example is a perfect one for a situation where two models might come into the picture together. In that case, the user commands the car to reach a destination which is impossible unless the car drives at a very high pace. If the car knocks down a person in this process, the liability would be part because of the user who gave such an instruction which would be covered by the natural-probable consequence model discussed above. Now consider another situation where the AI itself acts as the programmer of another AI program instructed to commit an offence. In such a situation, the third liability model is applied in addition to the first model. Thus, the three models of liability create a system of liability in which there is no escape from punishment.

However, there is another fundamental question that may arise with regard to AI entities. In case of a person, there is a certain standard of reasonableness which is expected of each human being.[40] Otherwise in all cases, law does not pay regard to the different cognitive capabilities of each individual. Special considerations are made for the sections of society which are not at par with the standard reasonableness associated with a normal human being. The exemptions given to minors (*doli incapax*) and the insane are examples of it. Similarly, each AI entity is not designed in the same manner; some might have superior capabilities than others. Therefore, some exemption is to be given to machines that cannot form an

[40]Regina v. Smith, 4 A.E.R. 289 (2000).

intention for an offence. How do we determine the capability of a machine and how do we set a parameter? The next part of this paper shall analyse this question in detail.

III. ASSESSING AI SENTIENCE: SEPARATING THE TRULY INTELLIGENT MACHINES

Although there is no doubt as to the prowess of modern day thinking machines, incriminating them is a different ball-game. The courts will have to determine whether it is worth punishing the AI entity and whether the entity fulfils the essential requirements of a crime. For this process of segregating actually intelligent machines from the non-intelligent ones, the courts will also have to employ a mechanism which serves this purpose much like when the courts take help of the bone marrow test while determining the age of the accused during a trial.

An intelligent entity is expected to possess certain attributes which are:[41]

(i) Communication: Communication is the most essential attribute while deciding the intelligence of an entity and therefore is the primary basis of the Turing test. The purpose is to look at the entity's spontaneous responses and determine how indistinguishable they are from human responses in similar situations.

(ii) Mental knowledge: An intelligent entity is expected to be aware of what it is and what it does, known in common parlance as sentience.

[41]Roger C. Schank, *What is AI, Anyway?*, THE FOUNDATIONS OF ARTIFICIAL INTELLIGENCE 3 (Derek Partridge & Yorick Wilks eds. 2006).

(iii) External knowledge: An intelligent entity must know and be aware of the external world and utilise that information while performing various tasks.

(iv) Goal-driven behaviour: The most basic purpose behind designing such entities, an intelligent entity is expected to exhibit and perform actions towards its task.

(v) Creativity: The benchmark of distinction between intelligent and non-intelligent entities. Certain level of creativity is required in all intelligent entities. For example, when a robot is asked to enter through door 'A' and it is unable to do so, it should take door 'B'.

Most modern day machines possess all of the above attributes.[42] However, the courts will still need to have a mechanism in case questions are raised over the intelligence of a machine. Intelligence in humans is typically ascertained by conducting exams. In case of AI entities, the examination most used is the 'Turing Test'. The Turing test was developed by British scientist Alan Turing in order to judge a machine's ability to exhibit behaviour that was indistinguishable to that of humans.

Under this test, a machine and a human interact through a text based interface and an evaluator marks the participants' responses on the basis of how much they resemble reactions that a human would give. If the evaluator fails to distinguish the reactions given by a machine and a human, the machine is said to have passed the test.[43]

This test however, has been subject to a lot of criticism. It was not originally devised by Turing to test the intelligence of machines but to prove that thinking machines could indeed exist. Also, the judgement of the evaluator has also been put to question. Another problem with the test is that it puts on the same pedestal consciousness and the

[42] *Id.*
[43] Alan Turing, *Computing Machinery and Intelligence*, 460 MIND (1950).

ability to simulate consciousness. Some human behaviour, it has been argued is unintelligent.[44] Although the result of the test may not be very precise, it certainly is correct enough to tell us if, or not, a machine is intelligent. Many have also raised concerns over how the application of this test will be done by the court. The various scientific tests already conducted are the answer to this question. The court merely has to order for a test to be done, and it would be carried out by an expert team. John Searle criticized the Turing test theory by his Chinese box test. He compared the Turing test to a man locked in a room who has no knowledge of the Chinese language and is given instructions in the language. But when he is given a rule book that consists translation of Chinese into symbols, he is able to understand the instructions. This convinces people outside the room that he understands Chinese, although the reality is quite different. John Searle thus puts forth the point that the machines act on the basis of algorithms and programmes which are manipulated on the basis of the inputs.[45] The Artificial intelligence entity actually does not think or formulate intention but acts on the basis of programs that function on the given input. However, even this theory has been criticised stating that although the program acts as a guide for the AI entity, it does not provide the machine sentience.

IV. PUNISHMENTS

Some scholars have expressed concerns over the nature of punishments for AI entities. The situation is further made difficult by the fact that modern day punishments are more inclined towards reformist forms rather than the retributive forms of punishment. The punishments given for crimes the world over majorly include capital

[44]*Artificial Stupidity*, 324 (7770) THE ECONOMIST (1992).
[45]Peter Kugel, *The Chinese Room is a Trick*, COMPUTER SCIENCE DEPARTMENT BOSTON COLLEGE, CHESTNUT HILL, USA, http://www.cs.bc.edu/~kugel/Publications/Searle%206.pdf.

punishment, incarceration and fines. In several countries there is also a prevalence of community service and probations. The question now arises whether the AI entities can be subjected to the same forms of punishment as humans are and to what extent will the effect of such punishment remain while punishing such entities. Similar questions arose when corporations were to be made liable and the answer to that was that similar punishments could be imposed on corporations as humans. For punishments which the corporations could not be subject to, changes were made, as required.

There are a few things that should be considered while punishments are decided. These considerations are applied in a similar manner and are comprised of three stages.[46] Each stage may be explained by a question, as described below:

(1) What is the fundamental significance of the specific punishment for a human?

(2) How does that punishment affect AI entities?

(3) What practical punishments may achieve the same significance when imposed on AI entities?

Let us discuss these questions in context of various prevalent punishments:

(1) Capital Punishment: Capital punishment is the most debated form of punishment all over the world. But the fact remains that it is one of the most effective deterrent in curbing crimes. For humans, it does one simple function, incapacitate the offender from committing the crime again. The same can be imposed on AI entities. Physical AI entities such as robots can be dismantled and their parts destroyed which could act as a deterrent, not for other robots but certainly for the programmers who make such devices. For software based AI as

[46]Hallevy, *supra* note 7.

well, the deletion of the software can act as effective incapacitation of the AI to commit further crimes.

(2) Imprisonment: Imprisonment is by far one of the most used forms of punishment. The main purpose behind imprisonment is that the offender gets time away from the society to prevent him from committing further crimes. Thus, for AI entities, although only physical ones, imprisonment can act as an effective measure to prevent them from committing further crimes. Although it may not have the usual characteristic of reform, it mays still work for highly intelligent entities that are capable to learn.

(3) Probation: An alternative to imprisonment is probation. Used majorly in Western jurisdictions in the form of suspended sentencing, the main aim of probation is to deter a person from committing offences by putting him in fear of punishment. Supervision is ordered by a court instead of ordering an imprisonment. AI entities can be kept under supervision after a first offence and a 'leash' can be put on them by their masters, or a tweak in their code may be ordered by the courts. They may also set a standard upon the autonomy of these entities.

(4) Community Service: Another widely used form of punishment, community service is mainly used in western societies. It is usually coupled with probation so that the offender pays the price for committing an offence. Considering the great functions that AI entities are capable of performing, they can certainly be used for community service. This will serve the dual purpose that a community service usually fulfils.

V. SCENARIO UNDER THE INDIAN LEGAL SYSTEM

The robotics industry in India is rising at a steady pace although it has still not reached the proportion that it has in countries like the U.S. or Japan.[47] Also, there have not been any reports of a serious threat or an incident where robots have played a major part in the crime in the country apart from the killing of a man in Haryana while working in a factory.[48]

Under the Indian Penal Code, all penalising provisions mention either "whoever" or "any person" or "any man". The definition of person under the Code does not actually define 'person'. It is merely an expansive definition of the term 'person' which talks about what person includes. It reads as: *"The word "person" includes any Company or Association or body of persons, whether incorporated or not."*[49]

So, if we were to bring robots under the scheme of liability of the Indian Penal Code, it could be done by amending the definition of the term 'person' under the code. The part to be inserted should be: *"it also includes artificially intelligent entities, whether physical or in the form of software, which have successfully passed the Turing test."* Such a definition would be quite comprehensive and would include all AI entities, whether physical or software based, it is also very precise in that it specifically includes only those entities which have passed the Turing test. Such a definition would clear any ambiguity on which entities would come under the provisions of the Penal Code. Amending this definition is also the way to go because it would not create any impact on the existing system of liability under the act. It

[47]R. John Edwards, *Robotics in India Starts Small but Is Growing Fast*, ROBOTICS BUSINESS REVIEW (Mar. 15, 2016), http://www.roboticsbusinessreview.com/article/robotics_in_india_starts_small_but_is_growing_fast.
[48]R.J. Singh & S. Yadav, *supra* note 2.
[49]The Indian Penal Code, 1860, No. 45, Acts of Parliament, 1860 (India), § 11.

would thus be very similar to the inclusion of companies and association under the definition of 'person'.

The amended definition would thus read as:

> *"11. "Person".—the word "person" includes any Company or Association or body of persons, whether incorporated or not. It also includes artificially intelligent entities, whether physical or in the form of software, which have successfully passed the Turing test."*

Another effective way of bringing these entities under the purview of law could be passing a separate legislation, a complete code for robots which would mandate registration of all robots irrespective of their capabilities. This code should also provide for the crimes and punishments which the robots commit. The liability under this code could be ascertained from the three approaches provided above. As discussed earlier, these approaches should be read conjunctively and liability be thus determined.

VI. CONCLUSION

Robots are the machines of the future. They are used in very diverse fields today and it is ever increasing. They are taking away places of humans in industries and defeating players of intellect such as Lee Se-Dol. Powered by AI, modern day robots can drive cars and answer questions on an intellectual level at par with the humans. Apple's Siri is an example of that. This is a clear example of their capabilities and what we are to see in the future. Therefore, there is no doubt that they can pose a threat to the human race.

Criminal law was devised to play a very important role in the society. Traditionally, criminal law never covered any species other than humans as it was deemed that no other specie could fulfil the

requirement of mens rea. This was the reason animals were never considered under criminal law and we only held their masters accountable for any untoward incidents. Robots were also excused from the ambit of criminal law as they were thought to function as per the whims of their masters. However, as technology progressed, so did the cognitive capabilities of the robot. The advent of Artificial Intelligence has made these machines 'think' and has breathed life into robots. The mens rea part of criminal law has peculiar requirements of intention and knowledge. The modern AI entities are so designed that they can have sensory reception of information through various devices embedded in them. Moreover, they are also capable of thinking and finding alternate ways to complete their tasks. Thus, it can be said that these modern thinking machines do fulfil the requirement of mens rea. Thus, it poses a very serious question that if these entities are capable of doing so much, then is it plausible to still punish the humans who program or use them just because we don't have the existing legal framework to punish them?

Corporate entities were once in question as to their inclusion under criminal law. They were deemed incapable of being subject to law as they could not have fulfilled the elements of crime, internal and external. However, it was in 1635 that an English judge first imposed liability on corporations which was a commendable and daring step. The law was such amended and devised that they could be brought under the ambit of criminal law. The same fear and anxiety that the society had at that time for corporations engulfs people today. The question of how to punish such entities and what are the feasible punishments for the same is of prime importance. The acquittal of the Swiss bot which purchased ecstasy pills because of lack of any legal provisions for the same is corroborative of the same.

Thus, considering the three models of liability provided above, we can use them together to create an inescapable net from which no entity can evade liability. Moreover, the approach should not be

construed so as to absolve humans of all liability. The only proposition is that thinking machines should not escape punishment. Also, the number of deaths caused by robots has been steadily rising in the past few years and the time is apt to make changes in our legal system and make robots liable for their acts, to set a precedent as was done by the English court in 1635. It would be appropriate to bring these intelligent machines under the purview of criminal law before they turn into *Frankenstein*'s robot.