

CRIMINAL LIABILITY OF ARTIFICIAL INTELLIGENCE~Prakruti N¹**I. INTRODUCTION**

In Austin, Texas, a group of protestors handed out “stop the robots” shirts in March 2015.² While it’s no doubt that intelligent robots are extremely useful, these protestors also believed that they could be extremely dangerous and further scientific development in the field should be done with extreme caution with specific focus on ensuring that these developments should be done safely.

In the 21st century, technology seems to be developing in an exponential rate. This development is so rapid that older generations often find themselves unable to catch up with the latest technology. Along with them, our laws and policies find themselves outdated or insufficient from time to time.

In 1969 the ARPANET network was invented, which was the first network to run on the TCP/IP protocol suite.³ In 1989 the World Wide Web (WWW) was born⁴. It took another decade after this for India to introduce the Information Technology Act.

AI is the capability of a machine to mimic intelligent human behavior. This allows the bot to function autonomously. While acting autonomously, a robot could cause injury or commit an act which would be criminal if committed by a person. In such cases, who is liable for such an act?

The Indian Penal Code defines “Person” as- The word “person” includes any Company or Association or body of persons, whether incorporated or not. It does not recognize AI as a person, and further has no provisions to determine who is liable for the actions of AI.

This research paper contains some models of determining liability from all over the world and analyses each one of them with respect to its relevance and applicability in India.

¹ The author is in her 3rd year studying at Maharashtra National Law University, Mumbai.

² Miller, R. (2015) Anti-robot protest held at SXSW, TechCrunch, <https://techcrunch.com/2015/03/14/anti-robot-protest-held-at-sxsw/>

³ Cox, K. A brief history of network technology, trueCABLE. Available at: <https://www.truecable.com/blogs/cable-academy/a-brief-history-of-network-technology>

⁴ History of the web (no date) World Wide Web Foundation. Available at: <https://webfoundation.org/about/vision/history-of-the-web/>

Criminal Liability Of Artificial Intelligence

Hallevy Models

Gabriel Hallevy⁵ has identified three liability models for offenses committed by AI.

1. The Perpetration-by-Another Liability Model

This is the simplest liability model. Primitive AI machines have decision making capabilities, but such capabilities are confined to its algorithm. These algorithms are straightforward to the extent that it has no black-box⁶ effect.⁷ In these cases, an AI is simply viewed as an innocent-agent.

However, it is important to note that in this model, an AI system is not simply equated with a non-intelligent machine, but rather equated to a mentally limited person, such as a child. This model is compatible with the existing provisions of the IPC.

According to S.108 of the IPC⁸, an abettor is a person who abets either the commission of an offence, or *the commission of an act which would be an offence, if committed by a person capable by law of committing an offence* with the same intention or knowledge as that of the abettor.

The definition of a person is not limited to, but inclusive of an association, company or body of persons. The distinction between simple robots and AI robots is drawn using the fact that AI robots have the capacity to mimic human intelligence and take independent decisions. This could potentially be enough grounds to also classify them as a 'legal person'. Such an inclusion is neither ground breaking, nor bizarre in light of the fact that Sophie, an AI robot was given the Saudi Arabian citizenship.

Further, in explanation 3 it is stated that "It is not necessary that the person abetted should be capable by law of committing an offence, or that he should have the same guilty intention or knowledge as that of the abettor, *or any guilty intention or knowledge.*"⁹ An AI robot does not need to possess mens rea, or even have the capacity to possess such mens rea

⁵ Hallevy, P.G. (2019) "The basic models of criminal liability of AI systems and Outer Circles," *SSRN Electronic Journal* [Preprint]. Available at: <https://doi.org/10.2139/ssrn.3402527>.

⁶ A 'black box' is a device, system or program that allows you to see the input and output, but gives no view of the processes and workings between. The AI black box, then, refers to the fact that with most AI-based tools, we don't know how they do what they do

⁷ Reed, N. (2021) *The Ai Black Box Problem*, *ThinkAutomation*. Available at: <https://www.thinkautomation.com/bots-and-ai/the-ai-black-box-problem/>.

⁸ Indian Penal Code, § 108, Acts of Parliament, 1860 (India)

⁹ Supra, note 5.

in order to have been abetted in a crime.

In this liability model, either the programmer or the user can be made liable depending on the origin of the perpetration. If the programmer, in the code of the robot, programmed the robot to behave in such manner that a law would be broken, he would then be held liable (as an abettor or perpetrator). For example, if a programmer creates a robot to generate automated responses in the customer service section of a particular retail website and includes in its program certain responses that could amount to sexual harassment, he could be liable for abetment of sexual harassment. However, if the code was innocent but the user used the innocent code to instruct the robot to do something that could potentially violate a law, then he would be liable.

However, an AI robot is not always an innocent agent. Sometimes an AI does an act that is against the law or harmful to another person, even though it isn't specifically programmed to do so. In such cases, no other human being has mens rea or actively contributes to the actus reus. In these cases, the previous model will fail.

2. The Natural Probable Consequence Liability Model

When an AI robot does an act which it wasn't explicitly programmed to do, which in turn amounts to any crime under the IPC, the liability will depend on the foreseeability of the crime.

For example, if a pilot is flying a plane and decides to turn on automatic pilot mode, an AI entity will begin to control the plane, much like a human pilot. The pilot gets news of a sudden change in weather and realizes a storm is brewing on the path the AI entity was planning to take. The pilot then tries to change the course of the plane which the AI entity sees as a threat to the safety of the plane. As a response, it decides to eject the pilot from his seat, thereby killing him.

The plane was never specifically instructed to kill anyone, much less a pilot. Machine learning allows a robot to access a large amount of data and learn from it. This data is unfiltered and the entity can learn virtually anything without the supervision of a human. It is possible that no human being had the knowledge that the AI robot had the requisite knowledge to eject a pilot from the plane.

Criminal Liability Of Artificial Intelligence

However, the auto pilot function had not been specifically restricted from ejecting a human. In the natural course of the auto pilot's functions, it was likely that it learnt to protect the plane from external dangers by way of ejection, cutting off oxygen supply etc. Here, there is no intention of the programmer or user to kill the pilot, but the mere negligence in this model amounts to *mens rea*.

When an AI entity is created, the programmer and the user must be able to foresee all natural and probable consequences of the algorithm and ensure that these are protected against. The programmer in this case should have ensured that the AI doesn't harm a person, especially the pilot, much less cause his death. Machine learning can be infinite, so the onus of adding checks and balances lies on the programmer.

In context of the Indian legal system, vicarious liability regime can be expanded to be applied here. Under vicarious liability, a parent is vicariously liable for the actions of their child. At the same time, a master is liable for the acts of their servant. The programmer shares a parental role towards the AI entity while the user has a master-servant relationship with the AI entity.

Both the programmer and the user must constantly be mindful of the natural and probable consequences of their use of the AI entity.

II. DIRECT LIABILITY

In the first scenario, a robot was viewed as an innocent agent. However, in the second scenario the AI entity is not treated as a child or a mentally ill person. The only differentiating factor between AI and human is *mens rea*. However, not all crimes require intention. In these cases, if an AI entity violates a law and the elements of that crime are fulfilled, the defence that AI does not possess *mens rea* cannot be used. However, Hallevy, at this point, emphasises that liability of an AI entity is additive in nature and does not substitute the liability of the person.

There are other liability models that can be employed to determine liability.

III. PRODUCT LIABILITY

"Product liability" is the responsibility of a product manufacturer or product seller, to compensate for any harm caused to a consumer by a defective product manufactured or sold or by deficiency in services relating thereto.¹⁰

When Google, and then Volvo and other car manufacturers released their self-driving cars, the liability of car accidents mostly fell upon the manufacturers. This is because these AI systems were considered to be a product purchased by the user and the manufacturer would be liable for the malfunctioning of such product.

However, product liability is a very niche liability model and can only be applied for those AI entities with highly specific use cases. Further, it only deals with civil liability and compensation, not criminal liability.

IV. STRICT LIABILITY

In theory, it seems fair that *Mens Rea* should be taken into consideration while determining the liability of a crime committed by an AI entity, especially when *Mens Rea* is a requisite element of that crime. However, it is true that robots are incapable of possessing such mens rea. When a machine learning AI robot is purchased or created, the onus of preventing such a dangerous object from causing hurt to someone lies on the owner of the AI entity.

The elements of strict liability are:

- Dangerous thing:

Artificial Intelligence is unpredictable. Coupled with this, it can also be physically stronger than a human. It has better computing power than a human. Unfortunately, it is also prone to making mistakes and lacks the basic morals and ethics of a human. In Japan, Kenji Urada was killed by a robotic arm as it mistook his identity and failed to recognise that he wasn't an intruder. Such mistakes are not made by the average human being which makes robots specifically dangerous.

¹⁰ Consumer Protection Act, 2019, § 2, NO. 35, Acts of Parliament, 2019 (IN)

Criminal Liability Of Artificial Intelligence

- Unnatural Land Use

In the case above, the robot was stationed in a factory which isn't unnatural land use. Neither is using self-driving cars. However, if one was to purchase a huge AI gadget for their backyard which then attacked a neighbour thinking of said neighbour to be a threat, this kind of use could amount to unnatural land use.

- Escape

This escape need not necessarily be physical. As long as the escape means going out of the control of the owner, any act done by the robot, outside what the owner intended to make the robot do, can be construed as an escape. However, both these elements are far-fetched as in most cases robots are used for very natural land uses.

A European Commission report states that "A person operating a permissible technology that nevertheless carries an increased risk of harm to others, for example AI-driven robots in public spaces, should be subject to strict liability for damage resulting from its operation"¹¹ However, there is no similar legislation in India to introduce strict liability to AI.

Currently, AI is either treated like a product, or a dangerous animal, a child or an actual human being. However, AI itself is a vast subject with many different subsets of differing sophistications. While most primitive AI entities only act as innocent agents, more sophisticated machine learning AIs even have the capability of being perpetrators.

Robots learn from not just unfiltered data but also patterns. This means that they are not only capable of learning information, but also the human biases that are intertwined with it. Robots are only getting more and more sophisticated and it is high time that our laws recognize the same.

A company can be considered a person and held liable for the acts of itself. Though a company is not an actual person, the laws don't only individually hold the members of the company's board liable. In the same way, the definition of person can be expanded to include AI entities. Or, a new definition can be incorporated to do the same.

Robots are capable of causing dangerous accidents. Sometimes they are coded improperly which results in them doing a wrongful act. If such robots are recognized to be dangerous,

¹¹ *Liability for Artificial Intelligence - European Parliament.* Available at: https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/

holding them liable could allow the courts to put down these robots or prevent the manufacturer from making more of such dangerous robots. Robots can inherit the ugly biases of humanity. Microsoft's AI chatbot Tay interacted with people for less than 24 hours on twitter before tweeting questionable and offensive tweets. While Microsoft claimed that this chatbot was filtered before release, in less than a day it was evident that this bot had no (or not enough) filters.¹²

If this bot was held to be liable for hate speech, it could be possible for the courts to order the bot to be shut down and never sold again. Criminal liability of AI may have no retributive effect (yet) but it still has some preventative effect.

V. CONCLUSION

Artificial Intelligence is developing at a fast pace but our laws are under equipped to deal with these entities. Currently, our courts have no choice but to make do with existing statutes to determine the liability of AI. However, the IPC catch up to the existing technology soon and define a clear set of rules to deal with the liability of Artificial Intelligence.

As AI is an ever-evolving field, and what we find ourselves in is just the beginning, our lawmakers must also keep an eye on the future and draft the laws to accommodate further advancements in AI and possible legal challenges they might pose.

¹² Wendehorst, C. (2020) *Strict liability for AI and other emerging technologies*, De Gruyter. De Gruyter. Available at: <https://www.degruyter.com/document/doi/10.1515/jetl-2020-0140/html?lang=en>.