REDEFINING CRIMINAL LIABILITY FOR ARTIFICIAL INTELLIGENCE HIGHLIGHTING THE RISE OF AI

Aditya Bhardwaj, Chanakya National Law University, India

Anushka (Utsav) Nag Mandal, Chanakya National Law University, India

ABSTRACT

Up until now, humans held a monopoly on being able to rationalise and form adaptable threads of logic to understand and comprehend the world around them. The human form of intelligence was held to be the only exclusive form of intelligence that had a real effect on civilisation. It was considered a relevant candidate for regulating and attracting punishment for the purpose of disciplining to fulfill the objective of having a better civilisation.

That has changed now. There is a new player in the field, a new form of intelligence that humans have created in their own image. This form of intelligence can understand and can participate in the world around them and perform functions which have a real impact on the economy, its surroundings, and its stakeholders. Artificial intelligence differs from other forms of technology in general and digital technologies in particular, a capability for self-development and can learn from the data that it receives and collects from various sources. This form of intelligence can actively interpret data present around it, produce outputs, and perform operations that it is asked to perform with its understanding of that data.

Since these features help to justify the attribution of criminal sanctions on human intelligence and conduct, should the features also justify the attribution of criminal liability, in relevant scenarios, towards Artificially Intelligent Systems and their conduct as well? This is the primary question that this article shall aim to probe and answer.

Keywords: Criminal Liability, Artificially Intelligent Systems, Medical Negligence, Large Language Models.

I. INTRODUCTION

In an ever-changing world driven by technology, Artificial Intelligence (AI) has moved from our mere imaginations to an integral part of our daily lives. However, till now there is no precise definition of 'Artificial Intelligence'. In common parlance, one can say that despite being a machine it possesses the "ability to adapt or improvise according to the feedback it receives to solve problems and address situations. These might go beyond the predefined set of queries and instructions that the AI was programmed to begin with[1]"! In other words, these systems learn from the data collected from the surroundings that they operate in and shape their learning curve to do their work better. From data-driven chatbots to self-driven cars, Artificially Intelligent Systems are reshaping businesses and altering the fabric of society. As AI gains more ground and applicability in our economies and societies, the question of how to hold them and their actions accountable becomes more pertinent. From the invention of early machine algorithms to the creation of advanced neural networks, AI has achieved a remarkable milestone.

A. Gradual growth and evolution of Artificial Intelligence (AI)

Today, AI plays a role in various critical sectors such as healthcare, where algorithms involving AI are used to diagnose diseases such as cancer, diabetes, and heart-related diseases through image and pattern recognition. In agriculture, AI is utilised to monitor crop health, predict weather, and in automated machinery which helps in harvesting and, plant crops. It also enhances transportation as self-driving cars can analyse real-time data for navigation. AI can also help in optimising traffic flow and in reducing congestion through data analysis and predictive modelling- those mentioned above are essentially self-driven systems which learn and develop on their own from the data that they gather from their surroundings to improve and perform better.

In the past years, AI has been rapidly growing and evolving, and as a result, definitions for what qualifies as AI have also been changing constantly. Just a few years ago, in the 2010s, AI was primarily defined as an ability of machines to perform functions or tasks that would normally require the assistance of human intelligence. Most of these functions were usually

[1] Ankit Kumar Padhy & Amit Kumar Padhy, *Criminal Liability of the Artificial Intelligence Entities*, 8 Nirma Univ. L.J. [Issue 2] (2019).

basic pattern recognition and problem-solving tasks. Artificially Intelligent (AI) Systems were initially designed to have narrow scope and applicability. For example, an AI system which was supposed to only recognise visual stimuli and a system which was supposed to recognise voices, was supposed to perform only that dedicated function, nothing more. This is something known as the Weak-AI hypothesis which propagates AI specialising in performing special, unique functions rather than a host of functions in general. This hypothesis was particularly prevalent and effective during the 2010s.[2]

The definition of what qualifies as Artificial Intelligence is much different and more inclusive now and involves systems which can perform and coordinate a vast variety of functions across different niches. This concept is called General Intelligence and is based upon a method of training AI called unsupervised learning, in which data is fed to the machine or system without labels and inputs from the developer as to how to interpret the fed data. This enables the machine to uncover hidden patterns within the data and use the data to learn a variety of things independently, without any prior guidance from developers or human agents.[3] The Weak-AI hypothesis is losing its appeal with corporate entities and scientific researchers alike focusing more on developing AI agents that can perform or coordinate a variety of tasks instead of those focusing only on a specific task or function.[4]

B. Legal and criminal liability associated with the involvement of AI in society

The technological advancements mentioned above are not free of challenges. A crucial point is brought up during the incorporation of AI into several industries: Should AI systems that resemble human intellect be held criminally liable? Now with AI systems growing more powerful and independent, this is not just a theoretical but also a practical question. Imagine for example, that an AI-backed self-driving car gets into a fatal accident because of reckless driving. Who is liable for this accident, human agents like the company and the developers behind the AI, or the AI itself, or both? Who to draw compensation from? In a similar vein, is it medical negligence if an AI-powered surgical robot performs a procedure and makes a serious mistake that harms a patient? What about an AI trading algorithm that commits fraud

[2] Antonio Lieto et al., *The Role of Cognitive Architectures in General Artificial Intelligence*, 48 Cognitive Sys. Res. 1 (2018).
[3] Google Cloud, *What Is Unsupervised Learning?*, Google Cloud (2022).
[4] Pradeep K. Dwivedi et al., *Opinion Paper: "So What If ChatGPT Wrote It?" Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy*, 71 Int'l J. Info. Mgmt. 102642 (2023).

by manipulating the stock market? These situations demonstrate how vital it is to rethink the structure of criminal liability in the context of AI systems.

More important than the aspect of economic issues is the aspect of protecting human rights and civil liberties in democratic republics when it comes to the use of Artificial Intelligence and systems powered by it, to enforce laws and legal provisions. Problems involving the use of AI in judicial proceedings and during administration of the law are also considerable issues. Currently, most of the laws or legal codes which seek to address the methods by which the government is allowed to enforce the laws framed, fail to consider the possibility of these governments using AI to enforce laws and legal liability over citizens.[5]

To address these issues, it is important to analyse the fundamental principles of criminal liability and to examine how these can be implemented in AI. "*If we talk about the criminal liability, which we all know is penal in nature, because punishment is a predominant feature of criminal proceedings, it basically not only requires a culpable act-actus reus (an action) but also requires mental state-mens rea (guilty mind) of defendant. So, the fundamental principle of penal liability is actus non facit reum, nisi mens sit rea: the act itself is not criminal unless accompanied by a guilty mind"*[6]. In addition to a forbidden activity, there must be proof of purpose or carelessness for someone to be held criminally accountable. There are some of the difficulties faced while implementing these ideas in AI.

Despite their sophisticated capabilities, artificial intelligence (AI) systems lack human-like consciousness and intent[7] They do not really seem to have a "guilty mind," since their operations rely on algorithms and data inputs. Furthermore, the assessment of criminal liability to AI systems calls for a reassessment of the functions played by designers, producers, and consumers. Is it appropriate to hold AI developers responsible for the behaviours of their creations? If so, how much of the concerned liability is to be borne by the developers of the respective systems? To handle these challenges and ensure that justice is upheld while fostering innovation, the legal system must evolve to address new forms of intelligence, such as Artificial

[5] V.F. Lapshin, S.A. Korneev & R.V. Kilimbaev, *The Use of Artificial Intelligence in Criminal Law and Criminal Procedure Systems*, 1001 IOP Conf. Ser.: Mater. Sci. & Eng. 012144 (2020).
[6] Sadaf Fahim & G.S. Bajpai, *AI and Criminal Liability*, 1 Indian J. Artif. Intell. & L., no. 1, 2020.
[7] Harry Haroutioun Haladjian & Carlos Montemayor, *Artificial Consciousness and the Consciousness-Attention Dissociation*, 45 Consciousness & Cognition 210 (2016).

Intelligence, which shares many traits with human cognition. This adaptability will allow it to meet the unique demands of an advancing society.

The successive chapters shall elaborate upon the present nature of Artificial Intelligence, the systems and sections of the economy powered by it, the nature and point of formulating criminal liability and how these two respective fields could interact with each other. It will also talk about how the aforementioned can be brought to a nexus to appropriately address the concerns regarding the evolution of principles of criminal liability in an era in which AI is only becoming more able, versatile and relative in the context of their involvement in economies and human civilizations. The ideal solution in this case would be to encourage the growth of AI as a form of technology as a powerhouse of productivity for the upcoming generations, while also framing provisions which will allow the potential victims of AI-initiated attacks appropriate compensations and safeguards to reduce the likelihood of similar incidents in the future.

II. A Contextual Background on Artificial Intelligence

The exciting and fast-growing field of artificial intelligence (AI) seeks to build machines that can carry out tasks that normally require human intelligence. Learning, reasoning, problem-solving, perception, and language comprehension are some of these tasks.

Defining AI is not easy; in fact, there is no generally accepted definition for Artificial Intelligence. Numerous different ones are used, and this can easily lead to confusion[8]

Artificial Intelligence (AI) is a term originally coined by emeritus Stanford Professor John McCarthy in 1955, he defined AI as "the science and engineering of making intelligent machines."[9] In this definition of McCarthy, there is a lack of specificity, and it is too broad in nature, as the word 'intelligence' in the definition leaves room for interpretation as the definition does not clearly delineate what constitutes 'intelligent' behaviour. Additionally, this definition does not consider the complexities and ethical considerations involved in AI. This may indicate that there was little clarity, in that era, pertaining to what Artificial Intelligence was exactly supposed to be. The Turing Test is still considered a viable test to indicate the

[8] Haroon Sheikh, Corien Prins & Erik Schrijvers, *Artificial Intelligence: Definition and Background*, in Mission AI: The New System Technology 1 (Haroon Sheikh, Corien Prins & Erik Schrijvers eds., Springer Int'l Publ'g 2023).
[9] John McCarthy, *What Is Artificial Intelligence?* (Stanford Univ.7, Nov. 2007)

intelligence of an Artificially Intelligent (AI) System. The Turing Test, which was originally named the "Imitation Game", is a test that is supposed to assess the intelligence of an AI system. To put it shortly, it states that if a computer can trick a human into thinking that a certain response it provides is similar to a human response, that machine can be said to possess intelligence.[10] It was incredibly rare for computers to crack the Turing test back in 1955.[11]

A. History of development of AI

The years of 1973 and 1988 were considered as AI winter as the funding to AI ventures was reduced during these years due to unmet expectations and technological limitations.[12] Despite these difficulties, some major developments took place during these years. During the 2000s, AI made significant strides due to advancement in algorithms, availability of larger amount of data and better computing power. AI also developed rapidly in areas such as natural language processing, computer vision, and speech recognition. Companies like Google, IBM (International Business Machines), and Microsoft invested heavily in AI research, leading to breakthroughs in areas such as deep learning and expanding on the work on neural networks. These developments paved the way for the integration of AI into various industries, from healthcare to finance to transportation[13].

Currently AI has become an indispensable part of the lives of a major section of society globally. It is not just limited to a few sectors or businesses, since AI can perform a range of tasks and provide better productivity with lesser cost as compared to human employees, with better quality of work in most of the cases. AI is revolutionising the education industry; it is being used to customise educational content and personalising education based on the strength and weaknesses of the students. AI is being integrated in gadgets like Alexa, Siri; and is being used to manage schedules, operate home appliances, and simplify daily tasks. In a survey conducted by IBM in 2023, it was highlighted that *42% of IT professionals at large organizations report that they have actively deployed AI while an additional 40% are actively exploring using the technology.*[14] AI is getting involved in almost every sector, from healthcare,

[10] *What Is the Turing Test? (Definition, Examples, History)*, Built In.
[11] S. Grand, *The Year 2001 Bug: Whatever Happened to HAL?*, 14 IEEE Intelligent Sys. & Their Applications 73 (1999).
[12] John Bohannon, *Fears of an AI Pioneer*, 349 Science 252 (2015).
[13] Martin Janse van Rensburg, "The Evolution of Data and AI in the 2000s - Adaptive AI Venture" (Adaptive AI Venture, Mar. 2024).
[14] MultiVu-PR Newswire, *Data Suggests Growth in Enterprise Adoption of AI Is Due to Widespread Deployment by Early Adopters, But Barriers Keep 40% in the Exploration and Experimentation Phases* (PR Newswire).

automobiles, education, computer vision, finance, travel and transport entertainment to social media and gaming, to name a few.

B. Key characteristics of AI

Several key characteristics of AI that distinguish it from any other technologies. One of the key characteristics is adaptability**.** AI can learn and adapt from the existing data or the data that has been provided to it. It can experience and learn by itself without much human interference, which allows it to improve itself.[15]

AI can also interpret and understand within the *context* of even large samples of data in little time, which enables it to perform more accurate and precise operations when it comes to performing the concerned functions that it has been allocated.

The ability to innovate and understand complex problems and the context in which they exist in, is something which is exclusive to AI among all other forms of computing technology. AI excels at analysing complex problems and finding solutions which is comparatively tougher for the human mind to calculate, and it processes, identifies, and makes use of that same data more efficiently than human beings.[16]

C. How is AI different from other traditional technologies?

Traditional technologies like conventional computing solutions which include manual computing or even narrow Artificial Narrow Intelligence, perform tasks that are predefined and specialised for a particular niche. Modern AI, like those based upon the concepts of Artificial General Intelligence (AGI) has the capacity to simulate human intelligence which also enables it to perform tasks which involves learning, reasoning and problem solving. AI can learn from the data provided to it and can improve itself as time passes, by learning from its expanding pool of information. Specially AI which has been trained via the process of unsupervised learning can adapt to the latest information and environment and perform its tasks better. AI can automate processes too complicated for traditional programming, like pattern recognition, natural language interpretation, and decision-making. This is because AI can learn, and its

[15] Raia Hadsell et al., *Embracing Change: Continual Learning in Deep Neural Networks*, 24 Trends Cognitive Sci. 1028 (2020).
[16] Sathian Dananjayan & Gerard Marshall Raj, Artificial Intelligence During a Pandemic: The COVID-19 Example, 35 Int'l J. Health Plan. & Mgmt. 1260 (2020).

complexity sets it apart from any other traditional technologies. AI consists of millions of lines of codes and multiple layers of algorithms which enables them to perform tasks the versatility, which is beyond the capacity of traditional technologies. AI, particularly generative AI can create fresh content such as texts, images and music. It can even create the entire movie just with proper prompt, this unique ability to generate novel outputs is a major difference between traditional technology and the modern AI.

III. Theoretical Concepts Concerning Criminal Liability

In the words of Woodrow Wilson, the 28th President of the United States of America, "Law is that portion of the established habit and thought of mankind which has gained distinct and formal recognition in the shape of uniform rules backed by the authority and power of the government".[17]

According to Salmond, "the law is defined as a body of principles framed and applied by the State in the administration of justice."[18] Hence, considering these definitive principles of what is right and what is wrong in terms of conduct and processes for both, the society and the individual, it becomes the duty of the individuals to abide by these principles and a duty of the State to enforce these principles and punish those who do not abide by them.

However, all the theories of punishment and statutory sanctions, which shall be elaborated upon later in this section, are geared towards punishing human beings and human elements of crime and violation of these legal principles and not elements which involve this new entire form of intelligence, that humans have invented, which acts as a human agency. This challenges the very concept constituting the nature and definition of crime in its present form.

A. The present nature of crime

The present nature of crime involves two crucial elements, Actus Reus and Mens Rea.

- Actus Reus- Actus Reus is composed of the physical element of the crime. It is the act, whose commission or omission, that constitutes a certain crime, as required by the provisions of a

[17] All Answers ltd, 'The Supremacy of the Law' (Lawteacher.net, July 2024)
[18] Supra Note 6.

statute. It is the voluntary act (commission) or the failure to act (omission), that typically gives rise to a criminally sanctioned result.[19]

- Mens Rea- Mens Rea is composed of the mental element behind the concerned offense or crime. It is mental intent behind the concerned commission or omission, which must be proved beyond reasonable doubt to really constitute a crime.[20]

The elements of a crime and its integral are geared towards human cases and therefore have a human niche of applicability. The punishment and its theories are allotted according to the fulfilment of these two essential elements and are also geared towards catering to humanistic objectives. The laws are designed with an inherent assumption that crimes can only be committed or attempted by humans, which was the case when humans held the sole monopoly on being able to deliberate on targeted themes. That is no longer the case since AI has come into the picture.

AI can learn, develop and reorganise itself based on a constant influx of recent data from its operating environment, alike to biological systems of learning as present in humans. It can very well display elaborate reasoning and intent behind some particular action that it is allowed to take, and the intent and reasoning can be their own and not consciously developed or coded by an engineer.[21]

This is the exact reason why these two essential elements of crime, *Actus Reus* and *Mens Rea*, apply in the case of AI as well. Therefore, dealing with wrongful acts committed via the help of AI becomes even more difficult to deal with. Regardless of AI fulfilling all the criteria which constitute a valid crime, there are no valid theories of punishment which have the appropriately dealt with the crimes potentially committed via AI.

Understanding what the lacunae in the various theories of punishment would require separate discussion.

B. Theories of punishment and associated lacunae

The concept of punishment in the field of criminal justice is based upon the idea that the State,

[19] Legal Info. Inst., *Actus Reus*, Wex: US Law, LII / Legal Info. Inst.
[20] James Ju, *What Are the Elements of Crime?*, Thomson Reuters: Law Blog (Jan. 30, 2024).
[21] Supra Note 15

acting as the Sovereign has the authority to punish an individual or an agent committing a crime. This is because each crime is a violation of a principle enshrined by the State and is a crime committed against the public-at-large and is assumed to have social connotations.[22] These are the following theories of punishment that might be relevant when considering the aspect of criminal liability involving AI:

a) Deterrent theory of punishment

The deterrent theory of punishment believes that it is possible and feasible to deter individuals from committing a crime in the future. This is done by making an example out of criminals who have committed a crime previously by imposing penalties and punishing them.[23] It suggests that the very possibility of being punished for doing a certain wrong deters an individual from committing it. They are further discouraged by witnessing the existing offenders being punished. It also aims at deterring the offenders from becoming repeat offenders.[24] It has had limited success in perfectly accomplishing its goal, because most crimes are committed out of erratic and impulsive decisions which fail to consider logical concepts like deterrence based on possible future implications.[25]

The ways that deterrence theory of punishment can be used positively in relation to the concept of criminal liability and punishment concerning conducts performed via the agency of Artificially Intelligent systems is in assurance against minor accidents (applicable in case of negligence or third-party liability) or minor infractions of the law caused because of the conduct of AI. If developers and owners are held liable for wrongful acts caused by these systems, they may be more motivated to address system flaws promptly. This responsibility could also encourage them to adopt a regular practice of thoroughly checking for and fixing system issues.

The concept of deterring future infractions will not be beneficial in the long term for the economy and society, concerning an upcoming industry like AI. Criminal liability is usually attributed to individuals, both natural persons (humans) and legal persons (companies).[26]

[22] Ambrose Y.K. Lee, *Public Wrongs and the Criminal Law*, 9 Crim. L. & Phil. 155 (2015).
[23] Dieter Dölling et al., *Is Deterrence Effective? Results of a Meta-Analysis of Punishment*, 15 Eur. J. on Crim. Pol'y & Res. 201 (2009).
[24] Murat C. Mungan, *The Certainty Versus the Severity of Punishment, Repeat Offenders, and Stigmatization* (Aug. 11, 2016).
[25] Elham Foroozandeh, 'Impulsivity and Impairment in Cognitive Functions in Criminals' (21 June 2017)
[26] Steven R Morrison, 'Relational Criminal Liability' (1 January 2016)

Individual liability for major violations of the law will attract major punishment and penalties to the perpetrator of the concerned crime. If we are to assume that punishments to the respective crimes is proportional to their gravity, and if these punishments were indeed supposed to have a deterring effect, that effect would count as a negative effect on the prospects of future development of AI. Plus, the economic connotations of such a negative development would be disastrous for the productivity growth of the workforce which counts as an affected group in this scenario as well.

Therefore, to deliver an appropriate punishment when major crimes committed via AI, the retributive theory of punishment would prove to be more effective compared to the deterrent theory of punishment.

b) Retributive theory of punishment

The retributive theory of punishment aims at the concept of attaining retribution or revenge from the concerned offender. It is based on the idea that offenders should suffer for the severity of the harms that they have directly or indirectly caused and should be held liable for and the quantum of punishment is proportional to the gravity of the harm caused.[27] It is based on the Shakespearean idea of "an eye for an eye", and it is the State which exacts the punishment from the offender. The associated penalty does not always have to be as harsh as it sounds. They can also take the form of monetary or economic penalties. The key element in exacting penalties under this theory is the element of "proportionality". The punishment exacted from the offender must be proportional to the damage caused by them.[28] The duty to ensure that the punishment imposed is indeed proportionate rests on a strict adherence to the due process of law and the judicial systems in the concerned country.

The retributive theory of punishment can be used appropriately in case of major crimes that are prospectively committed with the help of AI. Major crimes committed with the help of AI agents for which the operating company can be held responsible may be subject to retributive punishment. Common forms of punishments include substantial fines reflecting in the financial profile and goodwill of the company, probations including restricting certain acts

[27] Göran Duus-Otterström, *Do Offenders Deserve Proportionate Punishments?*, 15 Crim. L. & Phil. 463 (2021).
[28] Hadar Dancig-Rosenberg & Netanel Dagan, *Retributarianism: A New Individualization of Punishment*, 13 Crim. L. & Phil. 129 (2019).

or domains of power of the company and restitution in relevant cases.[29]

However, retribution is a very humane idea and is individual in nature.[30] It is the existence of the expectation that humans are supposed to follow a basic social contract, the infraction of which leads other humans to vouch for retribution and revenge in the first place.[31] Hence, if retributive punishments were to be imposed in cases of offences which involve AI agents, the retribution would focus on the human minds behind the AI and the corporation operating the AI system more than the fault facilitated by flaws in the AI system itself. This is problematic because any unfair use of AI in societal contexts can attract blame on the developers and companies associated with it, under the umbrella allegation of 'gross negligence'. This heightens the risk of unjust mass-moral policing, potentially discouraging economic growth and the further evolution of AI.

Therefore, to justify grounds for prosecution against offences conducted either with the aid of or with the initiation of Artificially Intelligent Systems, there are two ways to go on about it-

- Offensive acts committed by or with the help of AI can either be treated as Acts of God and with no legal consequences applicable to these acts or to those involved in these acts. This is because AI, as a non-human entity, cannot be held morally or legally responsible for its actions. Therefore, any harmful outcomes resulting from AI's operations could be considered unforeseeable and uncontrollable, coming from no predetermined and completely random patterns, like natural disasters or "Acts of God."

- Or these acts can attract legal liability and punishment for any human or human agency closely related to these acts involving the use or initiation of AI, instead of being seen in the same light as natural disasters with no ethical connotations involved to the concerned acts. This is because humans and human agencies in these AI systems can make conscious decisions and are capable of foreseeing potential consequences, unlike natural disasters which are uncontrollable and unpredictable.

[29] Guangming Gong et al., *Punishment by Securities Regulators, Corporate Social Responsibility and the Cost of Debt*, 171 J. Bus. Ethics 337 (2021).
[30] Neil Vidmar, 'Retribution and Revenge' (1 April 2000).
[31] Marco Faillo, Stefania Ottone & Lorenzo Sacconi, *The Social Contract in the Laboratory: An Experimental Analysis of Self-Enforcing Impartial Agreements*, 163 Pub. Choice 225 (2015).

The succeeding chapter shall discuss the constantly evolving nexus between the involvement of Artificial Intelligence in the economy and human society and that of legal principles governing the rule of law attracting criminal liability.

IV. The Nexus Between AI and Criminal Liability

As asserted and established in the preceding two chapters, AI is a rapidly evolving and growing field, and it is difficult to decisively define AI in the form that it exists in today. The development of Artificial Intelligence is still in its nascent stage. This stage of infancy in its research and its scope of development and application is something which is predictive of a phase of fast and defining growth in the coming years.[32]

The growth and development of AI have been divided into three distinct stages:

- ***First Stage*** - Artificial Narrow Intelligence (ANI), This is a stage of development of AI in which the Artificially Intelligent system focuses on specialising in doing one task better than the remaining ones. This is a stage in which the system is no more than a tool which helps the humans do what they want to do with better quality of work at their convenience.[33] Example: Visual recognition technology focuses only on the specialized task of visual recognition and nothing more. It is supposed to be used in consonance with a human mind driving the AI.

- ***Second Stage*** - Artificial General Intelligence (AGI), This is a stage in the development of AI which a vast majority of nations and economies in the world are yet to achieve.[34] When this stage of development of AI is fulfilled, the concerned Artificially Intelligent system can perform the same tasks with almost the same diligence and dexterity as exhibited by a humans in general.[35]

- ***Third Stage*** - Artificial Super Intelligence (ASI), This stage of development of AI is still a hypothesis. It is a stage in which the Artificially Intelligent system surpasses the physical

[32] Yogesh K. Dwivedi et al., *Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy*, 57 Int'l J. Info. Mgmt. 101994 (2021).
[33] 'Can AI Be Evil: The Criminal Capacities of ANI | International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE)'
[34] Caiming Zhang & Yang Lu, *Study on Artificial Intelligence: The State of the Art and Future Prospects*, 23 J. Indus. Info. Integration 100224 (2021).
[35] Fei Dou et al., *Towards Artificial General Intelligence (AGI) on the Internet of Things (IoT): Opportunities and Challenges* (arXiv, Sept. 14, 2023).

and other limits which traditionally bind and limit human intelligence. It hence becomes more powerful and intelligent than humans, effectively taking the AI system out of the domain of control of any one human or a group of humans.[36]

In this structure of imaging the development of AI, the trend of development has always been towards making AI as able as or more able than humans. As distasteful as it sounds, the move towards better AI has never been about being another tool in the hands of humans to help them do what they used to do better. It has been about making a thinking machine which can be more intelligent than humans and have the monopoly on intelligence that humans and human organizations have, be less relevant to economies and societies. A replacement of sorts.

Hence, due to this ambiguous nature of AI as being neither fully human nor an economic tool, there needs to exist a special set of mechanisms which aim at appropriate and fair sharing of liability for prospective crimes committed, between AI and humans.

The option which looks the most promising and implementable in the current context of what Artificial Intelligence looks like, is assigning vicarious liability to humans behind the actions of the AI system and assume the AI system as an innocent agent.

A. Vicarious liability of humans

Traditionally, when a human who does not qualify as competent to contract, even to appropriately decipher the social contract, like those who are held to qualify under the parameters of lunacy by a court of law or someone who is a minor or an animal, is held to be innocent even if they happen to commit an act which constitutes an offence.[37] This is because they are held or assumed to be incapable of having the mental capacity to constitute *mens rea* or a guilty mind.[38] If a sane person who fulfils the criteria for him to be competent to contract, instructs an innocent person who is presumed to be devoid of the capability to form a coherent *mens rea*, to perform an act which will be held to be an offence, the liability for the concerned

[36] Bahman Zohuri, *Artificial Super Intelligence (ASI): The Evolution of AI Beyond Human Capacity*, 3 Current Trends Eng'g Sci. 1 (2023).
[37] Georgios Tsimploulis et al., *Schizophrenia and Criminal Responsibility: A Systematic Review*, 206 J. Nervous & Mental Disease 370 (2018).
[38] Elizabeth Nevins-Saunders, *Not Guilty as Charged: The Myth of Mens Rea for Defendants with Mental Retardation* (May 24, 2012).

crime shall be on the former, the sane person who acts as the instructor.[39] The involved innocent person will not be considered an accomplice to the crime and shall be able to claim the defence and immunity of insanity or innocence.

In comparison to the above-mentioned scenario, if the innocent agent is held to be an Artificially Intelligent system and the sane instructor is held to be a typical human, if we are to abide by the aforementioned theory, then even if the human uses the AI system to commit a crime, the liability and the punishment of the crime shall befall upon the human and not the AI system, because the AI system is held to be innocent and incapable of forming a comprehensible *mens rea*.

This holds true and is practical in the current context of the scenario surrounding Artificial Intelligence. In its current form, Artificial Intelligence is still in the stage of being Artificial Narrow Intelligence (ANI) and cannot appropriately qualify as an Artificial General Intelligence (AGI) or Artificial Super Intelligence (ASI) and hence, it is appropriate to designate AI systems in their present form of existence, which is mostly Artificial Narrow Intelligence (ANI), as innocent in nature and devoid of the capability to form a comprehensible *mens rea*.

a) Model which considers AI as an innocent agent or intermediary

This model assumes that the Artificially Intelligent system has originally not been programmed to consider or conduct evil actions but only good actions. If this AI system ends up committing any evil, it is either by pure accident or by the instructions of the human associated with initiating actions with the system.

This is essentially a convenient way out of the sticky situation of being compelled to assess the complex and constantly changing nature of intelligence or cognitive capability possessed by an AI system. If we recount the details of what constitutes a crime, the two essential elements of a crime are:

- *Actus Reus* – The physical element of a crime. It is the physical act by an individual or a group of individuals that is in contravention to the provisions and principles of law as prevalent

[39] Herman, *Application of Liability Principle in the National Criminal Law*, 66 J.L. Pol'y & Globalization 15 (2017).

in a society or framed by a formal government.[40] For example, in cases of murder by stabbing, the physical act of the perpetrator plunging the knife into the body of the victim, is the *Actus Reus* of the crime.

- *Mens Rea* – This is the mental element of a crime. It is the mental element which is necessary to initiate the physical act, or the Actus Reus of the crime by a human being. Empirically, it is the knowledge that the concerned act is a crime, the conscious intent to commit a deed which is deemed to be a violation of the provisions and principles of the law, recklessness, wilfulness to commit a crime, and others.[41] Typically, for humans, *Mens Rea* is an essential element of a crime without which any act which is committed does not qualify as having been committed in deliberate contravention of the law.

The attribution of *Actus Reus* to acts committed by AI agents is relatively easy since it is all about the physical acts committed in the due course of the commission of the crime, but the attribution and determination of Mens Rea to the AI system is comparatively difficult, considering that AI systems, in their current form do not possess the ability to form criminal intent which can be incriminating enough to form *Mens Rea*. This inability raises several problems associated with legal liability of AI systems. Will defences like the defence of insanity available to humans be available to a malfunctioning AI machine? Will defences like the defence of intoxication be available to AI machines affected by an electronic virus?

B. Revisiting the basic assumptions behind Mens Rea

The jurisprudence behind the inclusion of *Mens Rea* as an integral constituent of a criminal act is something which for some, is an unbased assumption and for others a commitment to the idea of personal freedom and personal autonomy of individuals and humankind.[42] It is the idea that a person is responsible for their own actions and it is the very capability of humankind to avail this free choice of conduct which makes them attract compliments or blame and shame based on the choices they make. This idea, in turn, is the basis of a society which commits to democratic values, and to the idea of self-determination.[43] Self-determination stems from the idea that an individual is the basic unit of a society and is inherently valuable.

[40] Larry E. Sullivan, *The SAGE Glossary of the Social and Behavioral Sciences* (SAGE Publications, Inc. 2009).
[41] Michael J. Allen & Ian Edwards, *Mens Rea*, Law Trove (2021).
[42] Claire Oakes Finkelstein, 'The Inefficiency of Mens Rea' (2000).
[43] Sanford H. Kadish, *The Decline of Innocence*, 26 Cambridge L.J. 273 (1968).

If this basic idea is replaced by an idea which sees individuals as manipulable, curable and predictable, the resulting society will not be an individualist one.

Considering this context, Artificial Intelligence in its current form or in any of its foreseeable forms cannot be considered in the equal vein as a human individual would be.[44] The nature of Artificial Intelligence is significantly and inherently different from humans and Human Intelligence. Artificial Intelligence or a system powered by it, cannot be considered an individual in the same sense as a human being. The most plausible justification for this statement is probably the fact that Artificial Intelligence has no set parameters in terms of the personality that they might possess. In its current form, to possess a personality, an AI system simply could be instructed by humans as to what personality they might have, or an alternative approach may be to tailor an AI system which adapts to the user of the system and fine-tunes itself according to how it gauges the personality of its user from the data it collects from the user.[45]

Humans and their minds on the other hand can be considered Level 2 chaos. The very fact that the Chaos Theory which focuses on understanding the flows and rhythms of what, without a closer look, can be considered as dynamic and unstable systems, can be applied to humans is what makes humans different from computers or AI systems.[46] This illustrates that humans are built to be a part of the world and everything that is around them and not simply an isolated bubble functioning within it. Human beings and their minds change in incomprehensible and unpredictable ways in relation to changes in the environment around them.

Computers on the other hand are built with the purpose of primarily being simulative systems which have the capability to turn these simulations and formulations performed on or with or by them into interactive forms on command. The Chaos Theory can be applied to the effects of the functions performed by them but not natively as is the case with humans. This is perhaps for the better or for the worse. Some may compliment this as computers being 'resilient' to their surroundings or some may call computers 'inconsiderate' to what is around

[44] Ricardo Baeza-Yates & Pablo Villoslada, *Human vs. Artificial Intelligence*, 2022 IEEE 4th Int'l Conf. on Cognitive Mach. Intelligence (CogMI) 1 (2022).
[45] Byunggu Yu and Junwhan Kim, 'Personality of AI' (arXiv, 3 December 2023).
[46] David Loye & Riane Eisler, *Chaos and Transformation: Implications of Nonequilibrium Theory for Social Science and Society*, 32 Behav. Sci. 53 (1987).

them.

The question about what is then the ideal way to delegate liability when it comes to crimes or wrongful acts committed with the help of or by AI systems remains. Hence, the following sub-chapter shall elaborate upon it.

C. Respondeat Superior

The doctrine of *Respondeat Superior* shall be the primary doctrine for the time being, which shall determine the sharing of liability in cases of involvement of AI in commission of crimes or offences. As the article mentions above, this is considering the current nature of AI and how the current populace uses AI in the present and envisions to use AI in the future.

The doctrine of Respondeat Superior states that "the principal shall be liable for the actions of the agent who he assigns".[47] This illustrates the concept of vicarious liability of humans in cases of AI agents committing crimes or offences. Under this concept, the human or the organization which is associated with the AI system is considered the principal or the master and the AI system which is being used to accomplish a certain task, or which is directly associated with the *Actus Reus* of the crime, is considered the AI agent.

If the doctrine of *Respondeat Superior* is considered, then in cases of crimes committed with the help of or by AI systems, the human principal who initiates the process of the commission of the crime via a command or a prompt, shall be considered answerable for the concerned wrongful act. For instance, in cases in which a perpetrator uses deep-fake technology to generate a false media of some individual without their consent and use this media for wrongful purposes, the onus of the crime lies on the individual who started the process of generation of the concerned media and also prospectively on the corporate entity, if there is one involved, which caters the concerned technology as a service.

In cases in which the involved AI system is autonomous in nature, like a self-driving vehicle, in cases of an avoidable accident or other avoidable acts which cause harm to others, occur because of negligence, the onus of the incident and the duty for the payment of due

[47] Richard W. Crockett & Julie A. Gilmere, *Retaliation: Agency Theory and Gaps in the Law*, 28 Pub. Pers. Mgmt. 39 (1999).

compensation to victims lies on the corporate entity which operates the self-driving cars, and caters this technology as a service.[48]

V. Associated Case Study – Potential Misuse of AI

An idea is best demonstrated by the formulation of a case study, since readers learn more from examples than from descriptions. Hence, the following is a case study on the misuse of deep-fake technology and formulations of criminal liability in cases like this.

A. Introduction

The concept of deepfakes originated around 2017 when a Reddit user named Deepfake started posting doctored videos using AI software. Deepfakes are AI-generated and doctored videos of individuals doing things that they never actually did or wanted to do.[49] The key elements is that they are made without the consent of the individual who is being featured on the content and they are intentionally difficult to distinguish from authentic videos which makes deepfakes a potential disaster for democracy, privacy and an informed public body and hence, national security as well.[50] The history of deepfakes is, therefore, basically tied to the gradual evolution of Artificial Intelligence and methods to train these AI systems, especially through machine learning and deep learning. Running all the way from early experiments in neural networks and picture manipulation, deepfakes resort to quite distinct steps starting with the collection of large sets of sample images or videos of the targeted person. The larger the sample set of images or videos, the better and more realistic the deepfake will be. The images, videos, and audio created using advanced AI systems and techniques and are highly realistic yet entirely fabricated which is a real challenge to trusted authentic content on the Internet and aids disinformation, more intentional than unintentional.[51]

Through the consumption of millions upon-millions images, video clips, or voice recordings these algorithms grow to model human faces with incredible resemblance. Thanks

[48] Melinda Florina Lohmann, *Liability Issues Concerning Self-Driving Vehicles*, 7 Eur. J. Risk Reg. 335 (2016).
[49] John Fletcher, *Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance*, 70 Theatre J. 455 (2018).
[50] Thanh Thi Nguyen et al., *Deep Learning for Deepfakes Creation and Detection: A Survey*, 223 Computer Vision & Image Understanding 103525 (2022).
[51] Cristian Vaccari & Andrew Chadwick, *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*, 2020.

to the increase in data and computing capabilities, this technology is developing rapidly.[52] Deepfakes makes use of tools that are relatively easy to access and are openly offered on platforms which are open to access to any Internet user for free.

B. The Rashmika Mandanna deepfake case

In this case, the concerned actress' face was used without her consent to generate a video via an AI deepfake tool, which caused grave harm to her reputation and career. The concerned video was generated by a person called Eemani Naveen to boost followers on one of his social media pages. The video showed Rashmika's face superimposed on the torso of British Indian influencer Zara Patel in a skimpy outfit as she entered the elevator. The case sparked a discussion, for the first time in Bollywood and the Indian media industry, about the disadvantages and potential negative impacts of using generative AI technologies like deepfakes without proper regulation on the societal level.

Delhi Police charged the accused of violating Sections 465 and 469 of the erstwhile Indian Penal Code. The sections stood respectively for forgery and damage of reputation. They have been replaced by Sections 336 in the Bhartiya Nyaya Sanhita in consort with Sections 66C for identity theft and 66E for breach of personal privacy in the Information Technology (IT) Act, 2000.

C. Analysis of the case

This case highlights the inappropriate use of the generative deep-fake technology by which a video was generated using the face of Rashmika Mandanna and made use of a pictorial representation of her face on a body that did not belong to her, without her prior consent, causing her suffer grave harm to her reputation. The speed with which the video spread online, in general and on social networking sites, in particular, underscores a serious enforcement challenge that needs better solutions and highlights the susceptibility of the populace to misinformation and how difficult it can be to detect what information is false and what is true. This incident emphasizes the common ethical questions surrounding AI and deepfake technologies and highlights the need for stronger regulations surrounding AI, as well as effective ethical standards.

[52] Hina Fatima Shahzad et al., *A Review of Image Processing Techniques for Deepfakes*, 22 Sensors 4556 (2022).

D. Pros and cons of deepfakes

Talking about the pros of deepfake AI technology, it is now being used to raise awareness about sensitive societal issues in an entertaining and eye-catching manner. Businesses and companies are using deepfakes for better brand building and more creative advertisement campaigns at a lower cost as compared to traditional methods.[53] Except this, deepfakes are also revolutionizing the news media and entertainment industry by enhancing the visual effects and saving lots of resources. Deepfakes can also be used to bring back a deceased actor to life on screen, with the consent of their living survivors or prior consent taken from the actor before their death, which allows filmmakers to complete projects or create new project featuring actors who are no longer alive.

Alongside all the advantages that deepfake technology offers, it also raises ethical as well as legal concerns because of the potential misuses of it. Deepfakes can be used to create very realistic false videos or photos of anyone which can cause potential loss to someone's reputation. Deepfakes can be used to create misinformation by featuring influential individuals in fake scenarios without the consent of the concerned individual. It can feature politicians without their consent, in sensitive political scenarios and potentially cause social unrest.[54] As deepfake creates very realistic photos and videos it can be used to impersonate individuals, facilitating identity theft and fraud which includes creating fake IDs, passports, or even conducting financial transactions.[55] Deepfakes if not used ethically will create a general atmosphere of distrust, where people may doubt the authenticity of genuine videos and audio recordings.[56]

E. Notable constitutional provisions violated by the cited case

- **Section-336 (1) of Bhartiya Nyaya Sanhita (BNS)**, successor to Section 463 of the IPC – The section prohibits and criminalizes forgery. The Section defines the nature of forgery as making a false document or a false electronic record causing partition or departing from the property or entering a contract or cause damage or injury. In the above case by making

[53] Lucas Whittaker, Kate Letheren & Rory Mulcahy, *The Rise of Deepfakes: A Conceptual Framework and Research Agenda for Marketing*, 29 Australasian Marketing J. 204 (2021).
[54] Yisroel Mirsky & Wenke Lee, *The Creation and Detection of Deepfakes: A Survey*, 54 ACM Computing Surveys 1 (2021).
[55] Jan Kietzmann et al., *Deepfakes: Trick or Treat?*, 63 Bus. Horizons 135 (2020).
[56] Supra Note 51

a forged deepfake video of Rashmika Mandanna without her consent and making it seem legitimate enough to make a public imprint upon the reputation of the actress, the accused person is guilty of committing forgery.

In like manner, deep-fake technology can be further used to foster misinformation and forge records in the formal record-keeping system of the government and that can lead to more dangerous and influential errors in society, government and private matters, compromising the legitimacy of the entire record-keeping system in the process.[57]

- **Section 66c of the Information and Technology Act, 2000** – It states that if someone fraudulently or dishonestly uses another person's electronic signature, password, or any other unique identification feature which is registered in the name of another person, they can be punished with up to three years of imprisonment and may also be liable for fine or both. In the above case, a unique identification feature of Rashmika Mandanna, which is the picture or graphical imprint of her face is used against her will and without her consent to formulate a scenario which is damaging for her and her reputation, and with a disproportionate benefit to the perpetrator, which is the benefit of attracting more audience on social media platforms owned by them.

- **Article 21 of the Indian Constitution –** It is a part of the golden triangle of Articles of the Constitution of India and is one of the articles enumerating a crucial Fundamental Right. Article 21 of the Indian Constitution elaborates on the Fundamental Right to life, personal liberty and privacy.[58] The Constitution of India believes that everyone who the Constitution of India has authority over, deserves to live with dignity and respect and appropriate privacy. In the given case, the privacy of Rashmika Mandanna is breached, and her picture is used without her consent to create a false image or representation of her which ends up damaging her reputation and prospects. Hence, this is a violation of her Fundamental Rights guaranteed under Article 21.

VI. Examining Already Existing AI-related Laws in The World

Comparative examination of laws on AI in different countries is important to understand

[57] Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security* (July 14, 2018).
[58] Menaka Guruswamy, *Justice K.S. Puttaswamy (Ret'd) & Anr. v. Union of India & Ors.*, 111 Am. J. Int'l L. 994 (2017).

diverse strategies of regulation under the influence of different settings of political, economic, and cultural environments. It is this comparison that can be useful to find out best practices and innovative solutions to the problems arising due to AI, on one hand and to highlight the flaws in existing regulations to create more extensive legislation by filling the lacunae in them in their pre-existing forms. Furthermore, examining a bit more closely the way other countries deal with issues such as accountability and confidentiality could lead to improved levels of morality and legality. Different regulatory systems impact innovation differently, and drawing comparisons supports decision-makers in creating equitable policies that protect the public interest without curtailing innovation.

A. AI-related laws in India

The AI scene in India is relatively very nascent as compared to many other countries, that's why the laws and regulations in India that regulate AI are also very few.

- **The Digital Personal Data Protection Act, 2023**: The Digital Personal Data Protection (DPDP) Act, 2023 was passed with the goal of ensuring that the personal data of Indians electronically stored by commercial service-catering companies are stored within the physical boundaries of the territory of India and are taken abroad only if taking the data out of the territory of India is found respecting the list of norms and parameters set by the Act.[59] This Act made it mandatory to obtain explicit consent from individuals before processing or using their data. Through the enactment of this Act, the government has also attempted to assure safeguarding of privacy of the individuals whose data can be harnessed on the Internet by providing strict guidelines for the data fiduciaries and the establishment of the proposed data-protecting board will help to ensure compliance and address grievances.

  New forms of AI like generative AI feed on personalized data to solve cognitive and relational problems that the user instructs it to solve and tasks that the user wants it to perform. Hence, it is very imperative to regulate the data that is being provided to these systems for training. This Act has implications for the data collected from Indians that is used to train AI models hosted by companies registered outside the territory of India and further use of specialized AI trained on this concerned data.

[59] Ashwini Kumar, *The Digital Personal Data Protection Bill 2022 in Contrast with the EU General Data Protection Regulation: A Comparative Analysis*, 5 Int'l J. for Multidisciplinary Res. (2023).

- **Information Technology Act, 2000:** There is no mention of Artificial Intelligence in the Information and Technology (IT) Act, 2000. Still, its provisions may guide AI development in India to a certain extent.[60] Section 43A of this act provides compensation for such loss or damage caused due to a breach of data because of negligence of the corporate body or whoever possesses, handles, and deals with sensitive data belonging to persons and what qualifies as personal data. It says that body corporate shall be liable for payment of the compensation in such cases of loss. Again, the question arises of who is responsible for the breach of data caused during or by use of AI systems? Even though the Act does not discuss the criminal liability of AI systems, it could be interpreted and used to gauge the true nature of compensations to be provided in case such AI-related leaks take place.

- **AI Advisory published by Ministry of Electronics and Information Technology (MeitY):** In the view of the increasing use of AI within India, and the relatively unregulated economic operating space they could leverage, the Government of India issued an AI advisory dated 15th March, 2024 bearing guidelines for the free yet fair usage of AI within the territory of India, and to avoid causing harm to Indians and non-Indians from acts involving AI in India. The advisory from the Ministry has stipulated certain limitations on the intermediaries and platforms for not allowing its user of AI models to host, display, upload, modify, store, update, or share any unlawful content as mentioned in rule 3(1)(B) of IT Act 2000. The strict connotations and language of the guideline might have considerable negative effect on online freedom of expression in the country leading intermediaries to over-comply with government takedown notices and trample legitimate expression.[61] Intermediaries are also instructed not to introduce bias or discriminatory software in their AI models which may compromise the integrity of the electoral process. It guides the mandate of intermediaries modifying audio, videos, and images to include permanent unique metadata in that content for identifying the creator.

B. AI-related laws in the US (The State of California)

California has been one of the first states in the US to introduce laws which seek to regulate the training and application of AI systems and attempt to ensure fairness in treatment and

[60] Sheshadri Chatterjee, *AI Strategy of India: Policy Framework, Adoption Challenges and Actions for Government*, 14 Transforming Gov't: People, Process & Pol'y 757 (2020).
[61] Rishabh Dara, *Intermediary Liability in India: Chilling Effects on Free Expression on the Internet* (2011).

due compensation to anyone who might be affected by wrongful acts perpetrated via the use of AI.[62]

- **S.B 1047 - Safe and Secure Innovation for Frontier Artificial Intelligence Models Bill**: This Bill is lined up for a final vote in the State Assembly in August this year and is set to lay down crucial guidelines. The text of the Bill would require companies creating large enough AI models (extent to be decided after due deliberation) to put in place testing procedures and systems to prevent and respond to "safety incidents". The Bill looks very futuristic, requiring developers of AI models to implement certain security measures, one of which is ensuring they can implement a full shutdown in the shortest time possible, if need be. They are also required to report any incidents involving the safety of AI models to the recently established Frontier Model Division within the Department of Technology. This Bill further mandates the third party to have an audit of the system starting from 1st January 2028.[63] Mainly this act will deal with the safety and compliance of AI models to be trained in the future and the ones in use in the present.

- **California Autonomous Vehicles Registrations Guidelines by Department of Motor Vehicles (DMV):** To guide autonomous or self-driving cars, California has a detailed established set of regulations for both manufacturers as well as users of self-driving cars. Under these regulations manufacturers first need to obtain a permit from the Department of Motor Vehicles (DMV) to test autonomous vehicles in both categories, one with a safety driver and another for driverless testing.[64] Manufacturers need to meet the prescribed safety standards including the ability to operate without a driver for driverless testing. The manufacturers also need to submit an annual disengagement report to the department detailing the instances where the human driver had to take control. These regulations aim to ensure the safety of people and promote innovation to integrate driverless cars into the system safely.

[62] Mika Viljanen & Henni Parviainen, *AI Applications and Regulation: Mapping the Regulatory Strata*, 3 Frontiers in Computer Sci. (2022).
[63] Markus Anderljung et al., *Frontier AI Regulation: Managing Emerging Risks to Public Safety* (arXiv, Nov. 7, 2023).
[64] Bernard C. Soriano et al., *Regulations for Testing Autonomous Vehicles in California*, in Road Vehicle Automation 2, 45 (Gereon Meyer & Sven Beiker eds., Springer Int'l Publ'g 2015).

C. AI laws in the European Union (EU)

Artificial Intelligence (AI) Act of the European Union: This Act came into force on 1st August 2023 and can be termed a landmark in the development of AI laws; this act is very comprehensive and futuristic as it is not going to include only one domain but will extend to all domains, for example, automated vehicles, deepfakes, or any possible threat from AI, etc.

The AI Act shall attempt ensure that AI developed and used in the EU are trustworthy, with safeguards to protect people's fundamental rights, even though it has its own shortcomings.[65] It clearly defined what AI is and established a risk-based approach where it classified AI systems into four risk levels: Minimal risk, Specific Transparency risk, High risk, and Unacceptable risk.[66]

According to these categories, Responsibilities, Accountabilities, and Limitations of the AI models vary from each other. The Act spells out that the content generated by AI should be labelled as AI-generated, and the users must be informed that they are interacting with an AI system so that they become aware of the nature and context of their conversation.[67] It provides clear information about the system to be given by the developers of those AI models that come under the category of high-risk AI. Further, it disallows AI systems that are manipulative and deceptive in nature and, therefore, have the potential to affect the choice of a person in an uninvited manner. The Act shall establish a common regulatory framework so that there can be consistent standards and practices for AI across all countries of the EU.

VII. Conclusion

Artificial Intelligence is a fast-growing field which is only foreseen to grow further and faster than ever before and the definition of what the populace envisions when they talk about Artificial Intelligence is set to be subject to the most vehement and radical changes every few years.[68] The current form of liability which follows the doctrine of *Respondeat Superior* is

[65] Nathalie A. Smuha et al., *How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act* (Aug. 5, 2021)
[66] Claudio Novelli et al., *AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act* (May 31, 2023).
[67] Martin Ebers et al., *The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)*, 4 J. 589 (2021).
[68] Jiaying Liu et al., *Artificial Intelligence in the 21st Century*, 6 IEEE Access (2018).

applicable to the stage that Artificial Intelligence and systems powered by it are at in the present.

The current stage that Artificial Intelligence is at, primarily qualifies as ANI (Artificial Narrow Intelligence).[69] However, the move towards Artificial General Intelligence (AGI) is accelerating with the release and increased popularity of Large Language Models (LLMs) which are responsible for running a majority of the groundbreaking and popular generative AI applications of today like ChatGPT and others.[70] The special thing about Large Language Models is their advanced skill with language generation and comprehension. Language is a crucial part of how humans think as social beings and if the goal of moving towards AGI is the formulation of a form of Artificial Intelligence which can think and perform and interprets conducts initiated by itself or by others like a human, then it stands to reason that Large Language Models will be a key building block in the making of Artificial General Intelligence (AGI).[71]

The same legal principle or doctrine of *Respondeat Superior* which applies to ANI might not apply in the same sense and manner to AGI. AGI might not need a master to coordinate and orient its actions like ANI does and hence, the very premise of delegation of liability for wrongful actions committed via or by the Artificially Intelligent systems towards the master, developer or operators of the concerned systems will fall flat.

The delegation of liability under the doctrine of *Respondeat Superior* is justified by the prospect of the controllability and the preventability of the wrongful action that is supposed to be committed by or in consort with the AI system and hence, the very fact of the wrongful act happening makes the developer or operator liable under the prospect of negligence and lack of due care. The reason why this same justification will not work in case of AGI is because AGI is supposed to be able to coordinate multiple forms of ANI via a common medium or matrix.[72] That medium or matrix currently seems likely to be language and the ability to interpret and translate information into it. This ability of being able to interpret and act without human

[69] Ragnar Fjelland, *Why General Artificial Intelligence Will Not Be Realized*, 7 Humanities & Soc. Sci. Comm. 1 (2020).
[70] **Blaise Agüera y Arcas, Do Large Language Models Understand Us?, 151 Daedalus 183 (2022).**
[71] Bo Xu & Mu-ming Poo, *Large Language Models and Brain-Inspired General Intelligence*, 10 Nat'l Sci. Rev. 267 (2023).
[72] **Cong Guan et al., *Efficient Human-AI Coordination via Preparatory Language-Based Convention* (arXiv preprint, Nov. 1, 2023)**

prompting all throughout makes the process of delegating the liability of the AI to the operator or master considerably more difficult.

Hence, this article does not claim to be exhaustive in any way. What it is, is a comprehensive understanding of the modern context of development and growth of Artificial Intelligence as a technology and as an economic and social field and the interaction between this new economic and social field with the legal system. The coverage of the interaction between these two huge systems, for better understanding and comprehension, is focused on trying to interpret and develop a framework for sharing or delegation of liability in a just and fair manner, in cases of prospective wrongful acts committed via or by Artificial Intelligence systems while discussing the jurisprudence behind ideas propagated for this framework the whole time.