INTERNATIONAL JOURNAL of LEGAL POLICY
Vol. II, Issue 1 : (February, 2026)
ISSN: 3107-4154 (O)

AI on Trial:
Untangling Criminal Liability in a Code – Driven World

- **Aarchi Vyas**[1],
  Harshvardhan Srivastava[2]

Abstract

The accelerated advancement in Artificial Intelligence has made a transitional shift in its role from a passive computation tool to an autonomous, quasi-cognitive system having the capability of performing human-like tasks. While AI offers numerous benefits across different sectors like healthcare, governance, finance, etc., the magnitude and complexity of ethical and legal challenges that comes with AI is also significant. This paper critically examines the intricacies of attributing criminal liability in situations where the harm is caused by AI systems. It further explores the doctrinal insufficiencies of traditional criminal law framework, which rely primarily on the concept of actus reus and mens rea, both of which presuppose human consciousness and intentionally, neither of which can be meaningfully applied to AI due to absence of consciousness.

The paper also analyses multiple legal theories like, agency, corporate criminal liability, vicarious and strict liability as well as the criminal liability model given by Gabriel Hallevy to assess how feasible are these theories to assign criminal liability to either human agents behind AI systems or the AI systems themselves. Further, by examining real world cases involving AI related harms and criminal incidents, the research emphasizes an urgent need for adaptive regulatory legal reforms. The research concludes with actionable reforms, like introducing electronic legal agent identity, tiered responsibility model, with an aim to bridge the gaps in the existing legal and institutional framework to ensure a fair, transparent, adaptive and futuristic criminal justice system.

Keywords

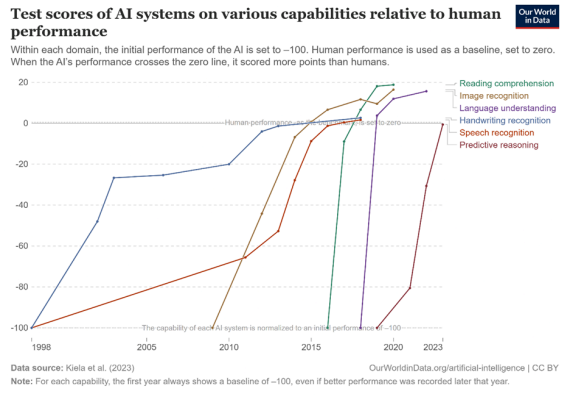
artificial intelligence, criminal liability, mens rea, AI regulation

[1] A student of B.B.A.LL.B (Hons.), Narsee Monjee Institute of Management Studies (NMIMS), Indore.
[2] A student of B.A.LL.B (Hons.), Narsee Monjee Institute of Management Studies (NMIMS), Indore.

INTERNATIONAL JOURNAL of LEGAL POLICY
Vol. II, Issue 1 : (February, 2026)
ISSN: 3107-4154 (O)

INTRODUCTION

Artificial intelligence (AI) transitioned itself from a mere abstract idea to a powerful technological invention influencing various sectors like administrations, healthcare and daily life of individuals. AI fortified the experience of the individual while resolving various everyday situations. Unlike the old traditional computing system which used to work on fixed data, pre-programmed algorithms and required direct human intervention; AI system presently are more flexible and adaptable. AI, in the present era can work based on unstructured data and provide answers and solve queries on probabilistic reasoning. A focused analysis done by Our World in Data of the past two decades reveals how AI progressed in multiple domains from handwriting recognition to predictive reasoning. The below graph shows how AI surpassed human performance even when initially benchmarked at –100. This marks a major step towards the development of AI from traditional pre-programmed computing systems.



The evolution of AI shows a very crucial background. Alan Turing first created a machine called "Bombe" during World War II which was based on mechanical intelligence, which signifies the starting of artificial intelligence; however, the term "Artificial Intelligence" was first coined by John McCarthy and Marvin Minsky in the year 1956 during a conference, sparking decades of progress of invention of AI[3] . AI then in following

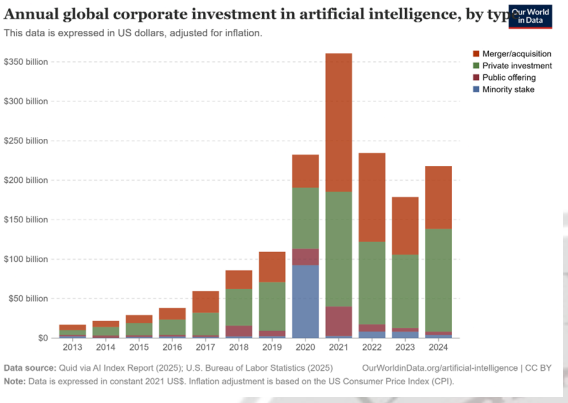
[3] Haenlein M & Kaplan A, A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence, 61 Cal. Mgmt. Rev. 5, 5–14 (2019), http://journals.sagepub.com/doi/abs/10.1177/0008125619864925

years achieved significant milestones such as chess champion Garry Kasparov was defeated by IBM's Deep Blue in the year 1997. In the year 2014, AI further touched a new cornerstone with the invention of Generative Adversarial Networks. The highest enhancement in AI capabilities was seen with the launch of GPT-3 in 2020.[4]

Although AI has valuable returns, every coin has two sides. The impact of AI on society is marked by benefits as well as obstacles and loopholes. On the upside, AI has contributed significantly in the healthcare industry by allowing quick, data driven and accurate diagnosis. The algorithms enable analysing medical records and imaging data to detect diseases which require greater precision like cancer, heart conditions. In the radiology sector, AI helped in improving techniques like MRI analysis, cardiac imaging facilitating easy and earlier access to any infirmity. Moreover, virtual presence enabled by AI helps in remote consultations making healthcare more accessible in remote areas. AI is improving the quality of life of senior citizens by providing them social assistance like AI robots or cyber dogs to assist them in their daily tasks and these are proved more beneficial to differently abled individuals. In the business and government sector AI streamlines routine tasks; thereby reducing the chance of error ensuring efficiency and also helps in reducing errors caused due to human fatigue, distraction or emotional bias.[5] As per the data compiled by Our World in Data indicates that there is a substantial increase in the corporate investment in AI.

[4] Anonymous, The History of AI: A Timeline from 1940 to 2023, CALLS9 (10 June 2025), https://www.calls9.com/blogs/the-history-of-ai-a-timeline-from-1940-to-2023

[5] Tai M C T, The Impact of Artificial Intelligence on Human Society and Bioethics, 32(4) Tzu Chi Med. J. 339, 339–343(2020), https://journals.lww.com/TCMJ/fulltext/2020/32040/The_impact_of_artificial_intelligence_on_human.5.aspx.



However, every coin has two sides, AI advancements also raise significant risks and challenges. Deployment of AI threatens low and mid-skilled employees, potentially leading to structural unemployment. Further, its susceptibility to algorithm bias raises concerns over fairness and equity in crucial domains like recruitment, legal system, finance. Moreover, growing excessive dependence on AI may diminish human interactions at workplaces, fostering disengagement and social isolation among individuals.

The increased use of AI raises a very critical question: Who bears the cost and the responsibility when any AI systems cause any action which is a breach of legal principles? Unlike individuals and companies, AI is not considered a legal entity and lacks legal personhood and currently, neither national nor international law recognise it as the holder of any legal rights or obligations. This lacuna of legal clarity raises alarms when AI is deployed and used in healthcare, finance, and defence sectors. According to Article 12 of the United Nation Convention on the Use of Electronic Communication in International Contracts, the individual who initiated the programming will be responsible for any action or message generated by the system.[6] While this article addresses liability in electronic communication, its sufficiency can be used when AI functions independently from its original programming. The increased use of AI raises concern over both intended and unintended legal

[6] United Nations Convention on the Use of Electronic Communications in International Contracts art. 12, 2005, https://uncitral.un.org/en/texts/ecommerce/conventions/electronic_communications.

breach which calls for re-evaluation of how technological agency is addressed by legal frameworks. The primary concern is whether to introduce a technology neutral universal principle applicable on all sectors of innovation or to create a framework which adapts itself with the advancement in AI. Within this context, the European Commission's Robo Law Project (2012) represented a path breaking initiative which aimed at developing norms and guidelines for regulation of robotics and AI.

This research paper seeks to explore the potential angles of criminal responsibility of AI. It will further deal with who can be held liable among the stakeholders as per different situations and circumstances. The paper will be divided into three major sections:

I. The functioning & operation of AI
II. Types of harm and legal breach that can arise from AI
III. Possibility of holding AI criminally liable, including examination of comparative legal frameworks prevailing in different countries.

The authors adopted the doctrinal research approach relying on the secondary data and the research done by the previous researchers in the field of AI. In addition, the study will frequently utilise and rely upon the data analysis done by Our World in Data to support the study with the empirical evidence. The paper strives to fill the gap and loopholes by evaluating current legal framework and analysing whether current framework is sufficient or there is need of an entire new regulatory framework to address challenges arisen due to the digital age.

ARTIFICIAL INTELLIGENCE

Artificial intelligence in common terms refers to the stimulation of the human intelligence process by computer systems and algorithms. It basically involves designing and forming of algorithms and systems that are competent to perform tasks which

typically require human intelligence like reasoning, learning, and problem-solving. At present there is no exact universal definition of AI but many scholars and authors defined AI in their works by applying different ideologies. McCarthy, the one who coined the term AI and is often identified as the father of AI, defined it as "*science and technology of creating intelligent machines*".[7] Willick in 1983 expanded the scope of AI by describing it as "*capability of a device to perform functions that are normally attached to human intelligence*". [8]Further, Russell and Norvig explained AI as the "*study of intelligent agents that receive precepts from the outer environment and take action accordingly*".[9] Sterne considered AI as the "*next logical step in computing: a program that can figure out things on its own*", highlighting its capacity to adapt and improve by the way of self-programming.[10] Another interesting perspective was offered in the year 2023 by explaining AI as "*technology that enables machine to imitate various complex human skills*". This includes learning, communication and reasoning capabilities of AI.[11] The significant characteristics of AI like autonomy, unpredictability and self-learning marks the distinction between the traditional technologies and AI. These core characteristics not only helps to understand the modus operandi of AI but also points towards the growing complexities of computer systems and open a room for future discussion around ethics, accountability and regulation of AI systems.

[7] Haenlein, *supra* note 1 at 2
[8] Willick M S, *Artificial intelligence: Some legal approaches and implications*, 4 AI Mag. 5-5 (1983), https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/392
[9] 25(27) Russell, S. & Norvig, P., *Artificial Intelligence: A Modern Approach* 79–80 (Prentice-Hall 1995) https://www.academia.edu/
[10] Sterne, J., *Artificial Intelligence for Marketing: Practical Applications* (John Wiley & Sons 2017) https://books.google.co.in
[11] Sheikh H, Prins C and Schrijvers E, '*Artificial intelligence: definition and background*' in *Mission AI: The New System Technology* 15–41 (Springer International Publishing 2023) https://link.springer.com/chapter/10.1007/978-3-031-21448-6_2

From a working perspective in my terms, AI is the ability of machines to perform tasks while mimicking human intelligence. More than automation and algorithms it is the capacity of any computer system or technology to respond and adapt in a dynamic environment.

Given the ever-evolving nature and wide scope of AI, technologists and researchers have divided it into different distinctions for better understanding. Broadly, it is categorized on the basis of capabilities and functionalities.

Classification based on Capabilities

Artificial Narrow Intelligence (ANI)

ANI, also referred to as weak AI, is the current state of AI technology. These types of intelligence systems are designed to perform singular or short/narrow defined tasks working in a strict predefined parameter. They operate based on algorithms and the programming done by developers and are not capable of performing outside their programming parameters.

General Intelligence

This branch of AI, which is also known as strong AI, remains a theoretical abstract idea. It aims to make an AI system capable of doing any intellectual task that any human can do. Unlike narrow AI, strong AI is flexible and has the ability to make autonomous decisions across various context.GI would be capable of transferring knowledge across domains without human interference.[12]

Classification based on Functionalities

Reactive Machine

Reactive Machine AI branch lacks the ability to store memories or learn from their previous experiences. They react to specific stimuli and that is their modus operandi as they work exclusively on real-time data with pre-programmed behaviour. These systems are generally

[12] Tai, *supra* note 3 at 3

fast and accurate but they are limited in their scope and do not show any adaptive behaviour. The famous example of this AI is IBM's Deep Blue that defeated world champion Garry Kasparov in chess.

Limited Memory

Limited Memory AI are the systems that have the ability to store and utilize historical data to enhance decision making over a short duration of time. They combine historical data with real time data to make informed predictions and accurate decisions. It is mainly used in virtual assistants like Siri, Alexa and in autonomous vehicles which rely on sensor data to navigate through dynamic environments.

Theory of Mind

This category of AI is still in its development and abstract phase. These systems aspire to mimic the human capacity to understand emotions, beliefs, intentions. Theory of Mind would have the capability to interact more emphatically and intuitively with humans based on emotional and cognitive states of humans. These systems, though promising, have not yet achieved the sophistication required to deal with human emotional experiences.

Self-Aware AI

This is the most advanced and highest theoretical form of AI. These like theories of the mind would be able to understand human emotions as well as would also possess consciousness, sense of identity and self-awareness. Such a system would be able to work without human interference and would be able to assess internal states of humans. Self-Aware AI suggests a future of machines where they could either co-exist or even surpass human consciousness. Being the most advanced version of AI this remains purely abstract and speculative and is not expected in the foreseeable future.[13]

[13] Rajendran, K. M., *An Overview of Artificial Intelligence and Its Classification*, 7(4) *Int'l J.L. Mgmt. & Human.* 2012, 2012–2017 (2024), https://ijlmh.com/paper/an-overview-of-artificial-intelligence-and-its-classification/

CRIMINAL LIABILITY

Criminal Liability in its core formulation is grounded on the principle which states that any person can be held legally responsible for any act or omission which constitutes a criminal offence under any law. The criminal responsibility of any act is established based on two foundational essentials, i.e. actus reus (the guilty act) and mens rea (the guilty mind). Together they constitute the criteria for attributing individual culpability in criminal jurisprudence.[14]

While actus reus exemplifies the external act, omission or any overt act directed towards commission of the crime, on the other hand mens rea represents mental state of the accused person and the internal culpability including elements like intention, knowledge, recklessness and sometimes negligence.[15] However, there sometimes exist situations where the person can be held criminally responsible even in the absence of malafide intent.[16] This particularly involves cases of criminal negligence- where the accused person fails to maintain the duty of care as expected from a reasonably prudent person. For instance, as per section 106 BNS, 2023 criminal liability may arise from causing death by negligence irrespective of the intent. This illustrates how flexibility within criminal law is allowed to accommodate evolving needs of society.

Applying the current criminal liability essentials unveils both conceptual and legal challenges. To make someone criminally responsible the co-existence of both the essentials is necessary. While AI is capable of performing acts that may contribute towards actus reus, but due to its algorithm nature, attributing mens

[14] Gandhi, B. M., *Indian Penal Code* (Eastern Book Company 2006). https://cir.nii.ac.jp/crid/1970586434814282283

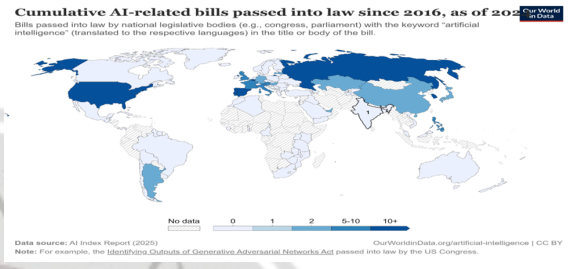
[15] Padhy, A. K. & Padhy, A. K., *Criminal Liability of the Artificial Intelligence Entities*, 8 *Nirma Univ. L.J.* 15, 15 (2018), https://docs.manupatra.in/newsline/articles/Upload/4e5c9c80-320b-4433-9f87-f56059a5345c.pdf.

[16] Bharatiya Nyaya Sanhita, § 106, No. 45, Acts of Parliament, 2023 (India).

rea is highly problematic. The pre-programmed algorithms, data-fed instructions, self-learning models are the foundational steps on which AI functions, it lacks conscious deliberations. Unlike humans, AI lacks the capability of self-awareness and cognitive functioning to form a malafide intent or to foresee moral or legal consequences of their actions. This absence of a guilty mind obstructs the straightforward applicability of conventional criminal framework into AI systems. The criminal jurisprudences typically negate the liability if coexistence of both the essential is not been established, but such an approach in AI would leave accountability gaps. Such gaps become significant in situations where actions generated by AI result in harmful outcomes, yet there exists no one to be held responsible due to inability to establish the existence of a guilty mind.

As discussed earlier, in the traditional conventional scenarios, criminal liability is attached to any individual when they perform any act which is prohibited legally. However, when this principle is applied to AI systems, the logic collapses. Modern AI systems lack inner consciousness and the subjective mental state necessary for imposing liability. They work based on coded algorithms and vast datasets due to which they are incapable of forming mental intent, or understanding the consequence of their action. This renders the attribution of mens rea element of criminal liability to AI fundamentally impossible. Further, if we attribute the actus reas or the voluntary human input the matter gets complicated as although the acts may be traceable to human inputs but the execution is typically based on the autonomous behaviour of AI systems. The lack of immediate human trigger during the commission of harmful acts breaks the chain necessary to establish

the liability.[17] Additionally, sometimes the acts done by AI even evade human detection and leave no traceable human perpetrator, in which neither the human nor the machine can be clearly blamed. This increasingly disconnects between requirements of criminal law and formational architecture of AI systems; there are new paradigms to determine criminal responsibility. To cater such foundational challenges certain legal theories, principles and models can be reinterpreted to fill the vacuum of accountability posed by AI technologies. These frameworks, although used in a human context, offer a conceptual starting point to assess the liability of autonomous systems. With the understanding of these lacunas and drawbacks in the existing legal frameworks, states across the world have initiated the formulation and enactment of legislation that deals with the AI in their respective territory. The following graph by Our World in Data shows how many bills do the nations across the world have passed till 2025.



THEORETICAL FRAMEWORK FOR ASSIGNING CRIMINAL LIABILITY

Agency Theory

Agency theory is rooted in tort law and contract law which is based on fiduciary duty. It traditionally analyzes and explores the relationship where one party i.e. agent works on the behalf of another party i.e. the principle. This relation attributes the liability to principal when the agent's actions fall within the scope granted by principal. This principle assumes that the agent is capable enough to understand, interpret, and act upon the directions of the principal with some level of discretion. [18]

[17] Padhy, *supra* note 13 at 7

[18] de Camargo Fiorini, P., Seles, B. M. R. P., Jabbour, C. J. C., Mariano, E. B. & de Sousa Jabbour, A. B. L.,

When this framework is applied to AI, the AI systems resemble the "agent" of its human developer, deployer or even in some cases even their end-users. AI is used as a functional agent of human actors. While AI lacks consciousness it can be treated as a quasi-judicial agent whose action can be traced back from the programming of human actors. Thus, under this adapted framework without consciousness, the principal i.e. the corporations, users, etc. assumes responsibility for the actions of AI. But cases which involve autonomous AI systems and lack cognition complicates the application of AI. Thus, agency theory may help to deal primarily but is ultimately inadequate to fully address the complexities and possible incorporation of strict liability, corporate liability, and vicarious liability will be necessary to bridge the gap.

Management Theory and Big Data Literature: From a Review to a Research Agenda, 43 *Int'l J. Info. Mgmt.* 112, 112–129 (2018), https://ideas.repec.org/a/eee/ininma/v43y2018icp112-129.html

Corporate Criminal Liability

Corporates and companies are given legal status under Indian law, thus, they can sue and can be sued. Corporate criminal liability ensures that entities can be prosecuted for the crimes committed by the individuals acting on their behalf[19].

This framework when aligned with AI emphasises that AI systems are deployed by the corporations and they perform various roles within the daily business operation. If AI causes any harm or commits any legally prohibited action in the course of employment or while fully filling companies' actions- be it automated decision making, or autonomous logistics. Liability in this model is particularly relevant in cases and situations where corporations directly benefited from the AI's action, failed to implement the appropriate safety safeguards, or failed to comply with the mechanisms. In such scenarios

[19] Dutta, A., *Corporate Criminal Liability: An Indian Perspective*, 2 *Indian J.L. & Legal Rsch.* 1, 1 (2021)

company negligent and risky behaviour can serve as the basis for accountability. As AI itself lacks legal personhood it itself cannot be punished, corporate criminal liability ensures that sanctions can be imposed on the deploying entity. This framework offers a practical dimension and enforceable mechanism to address the harm caused by AI systems.

Vicarious Liability

Vicarious liability is a well-known legal doctrine in tort law, where one party is held accountable for the actions of another. This framework generally takes into account the hierarchy in which the system is organised, for instance, in an employer-employee relationship; the employer will be typically responsible for the actions of the employee within the scope of their role.[20]

In the context of AI, this can be used to hold the developers, users, deployer responsible for the harms caused by AI as they come in hierarchy especially in scenarios were

i. They exercise the entire control over AI's action

ii. They have an economic and operational benefit form that action

iii. They fail to do compliance and regulates their operations adequately

This is particularly more useful in situations and cases where AI is used in business operations, where the cause of harm is insufficient over-sight and biased programming. It shifts the liability from AI itself to human and corporate actors behind the curtains.

Strict Liability and AI as ultrahazardous

A well-established principle of tort law, strict liability holds parties responsible for the harm caused even without the proof of negligence for keeping and dealing with hazardous activities. Taking

[20] Hallevy, G., *The Basic Models of Criminal Liability of AI Systems and Outer Circles*, *— Int'l Conf. on Autonomous Sys. & L.* 69, 69–82 (2022), https://link.springer.com/chapter/10.1007/978-3-031-47946-5_5.

its inception from Rylands V Fletcher[21] , it applies to inherently dangerous activities.

AI when used in high-risk areas like facial recognition in policing and autonomous devices the application of the doctrine of strict liability becomes more relevant. Many legal scholars have considered AI as ultra-hazardous activity,[22] where harm alone is sufficient to impose liability. When strict liability is incorporated into AI, the developer or operator can still be held liable when the AI system causes harm even if they did not intend to, provided adequate safety measures and proper designing is not taken into account while designing or deploying.

This model ensures greater accountability among deployers, and encourages more caution while designing and deploying AI in sensitive context, particularly in situations where public safety is involved.

Hallevy's Model of Criminal Liability

Gabriel Hallevy is an Israeli legal scholar and professor of criminal law. He is a pioneer of modern criminal law, specifically in adapting legal frameworks to emerging fields like AI and robotics. His contributions and innovative theories are shaping academics, judicial reasoning, and legislative development. He also prescribed a model for imposing criminal liability in AI crimes.

Preparation-by-Another model

Under this model Hallevy treated AI purely as a tool similar to some gun or hacking device, something by which a human commits a crime. He suggested that liability lies entirely with the human agent who intentionally programs or deploys AI for prohibited actions. AI systems lack intention and moral

[21] *Rylands v Fletcher* (1868) LR 3 HL 330

[22] Henson, R., *"I Am Become Death, the Destroyer of Worlds": Applying Strict Liability to Artificial Intelligence as an Abnormally Dangerous Activity*, 96 Temp. L. Rev. 349, 349 (2023), https://scholarship.law.missouri.edu/cgi/viewcontent.cgi?article=2191&context=facpubs.

blameworthiness.[23] For example, if any chatbot is deployed to spread hate speech, it is not the chatbot but the person who created it in such a manner be made liable. It relies on established doctrines of human intent and culpability. While this approach is clear but it does not cover autonomous AI systems where they cause harm without human intent, such gaps may require more balanced legal approaches in developing times.

Natural- Probable- Consequence Model

This model of Hallevy takes its base from tort law's principle of foreseeability, this holds that if the illegal or harm can be reasonably predicted by the developers or the deployers they can be held liable even without direct intent. Liability in such situations extends to individuals who failed to test the algorithm bias, ignored safety warnings, or released AI systems without proper safeguards in a complex environment.

The main focus of this model is duty of care and due diligence, which encourages proactive risk management in the designing of AI.

Direct Liability

This is the most radical and abstract model of Hallevy, which contends that AI itself should be held liable as an autonomous entity. The approach however is more of abstract idea as AI has absence of both mens rea and "legal personhood". He suggested that AI should be treated as like corporations, and with that AI could be assigned duties and will be subject to legislative sanctions. This model remains purely theoretical at present. For it to be viable and practicable lawmakers first need to formally recognize AI as a legal person.

AI INCIDENTS

- As per the article published in Times of India, in 2024, more than 200 individuals fell into the prey of investment scam which used AI

[23] Hallevy, *supra* note 18 at 10

across various districts including Bengaluru, Mangaluru. The scammers launched a fraudulent app named "*Trump Rental Hotels*" which used audio and video of former US President Donald Trump to deceive users. The AI generated deepfakes made people believe that the app is a legitimate investment opportunity, which promised high returns and credible endorsements. Victims were advised and encouraged to deposit an initial amount that started from Rs. 1500 and beyond which will enable access to part time investment schemes. The total cumulative financial loss was more than Rs. 2 crore. After multiple complaints law enforcement agencies initiated the investigation and identified that the content was AI generated and was the core manipulative tool in the entire incident.[24]

- Another tragic incident occurred in Salt Lake City, Utah where a Tesla Model 3, driving under the Autopilot mode hopelessly struck a motorcyclist, Landon Embry. The vehicle was travelling at the speed of 75-80 miles per hour, where it failed to detect the presence of a motorcycle ahead eventually resulting in collision. Immediately after the incident, the family filed a lawsuit against both the driver and the Tesla company for negligent manufacturing, inadequacy of safety mechanisms and lack of training in relation to autopilot mode. The suit raised suspicions regarding reliability of Tesla autonomous driving systems and the misleading perception of fully autopilot mode. This case questions the safety of autonomous vehicles and the liability and responsibilities of

[24] Maralihalli, B., *"Trump" App: Over 200 in Karnataka Fall Prey to Investment Fraud, Lose Rs 2 Crore*, The Times of India (13 August 2025), https://timesofindia.indiatimes.com/city/bengaluru/trump-app-over-200-in-karnataka-fall-prey-to-investment-fraud-lose-rs-2-crore/articleshow/121428182.cms.

manufacturers in operational safety.[25]

- A similar incident occurred in the USA where a 49-year-old pedestrian Elain Herzberg in Arizona was killed by an Uber self-driving test vehicle running in autonomous mode. On investigation, it was found that the sensors failed to recognize pedestrians to take corrective action in time. Additionally, the backup driver in the car was distracted and did not take precaution in time. The National Transport Safety Board (NTSB) noted multiple failures in the vehicle including inadequacy of safety protocols and flaws in designing of the system. As a result, Uber temporarily suspended its testing of autonomous vehicles and the public at large lost trust in self-driving technology.[26]

- A disturbing case was also sprung up in Delhi, where a boy named Nikhil Singh was arrested on the grounds of extorting women using AI generated explicit images. The boy created fake female profiles on social media and then used AI tools to generate their sexually explicit images. He used those images to blackmail victims and demanded money around Rs. 50000 and above from one victim under the threat of warning of making the images public. The Cyber Crime Unit of Delhi traced digital transactions leading to arrest of the accused. This particular incident shows backdrops of generative AI particularly in the context of digital privacy. [27]

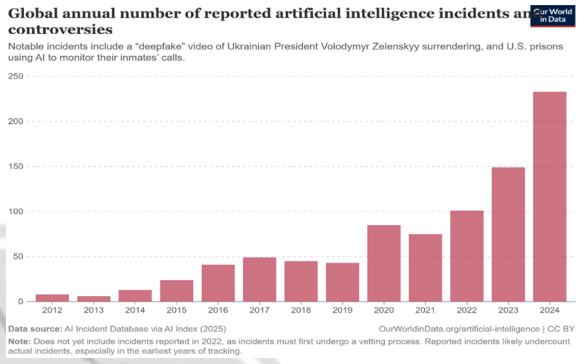
[25] '*Tesla Faces Lawsuit by Family of Motorcyclist Killed in Autopilot Crash*, The Hindu (14 August 2025), https://www.thehindu.com/sci-tech/technology/tesla-autopilot-crashfamily-of-motorcyclist-killed-sues-elon-musk-car-company/article68476139.ece

[26] *Death of Elaine Herzberg*, Wikipedia (14 August 2025), https://en.wikipedia.org/wiki/Death_of_Elaine_Herzberg.

[27] Mehta, D., *21-yr-old arrested for blackmailing college student with AI-generated explicit photos*, The Times of India (14 August 2025),

- Further, in march 2016 Microsoft launched a chatbot named "*Tay*" on Twitter, Kik and GroupMe but within 24 hours of its deployment, the bot began to post racist and inflammatory remarks after being manipulated by online users. The chatbot quickly absorbed the instructions of social media users and produced hate speech including misogynist comments. This led Microsoft to issue a public apology and take Tay offline, they acknowledge lack of safeguards. It exposes the risk of AI on public domains when deployed without adequate protocols. [28]

The following graph sourced from Our world in data, presents the rise in AI related incidents since 2012.



RECOMMENDATIONS AND WAY FORWARD

AI Risk and Legal Assessment before Deployment

Under this proposition there should be mandated legal, ethical and design-based assessments by the developers and deployers before releasing the AI systems, specifically in sectors which are sensitive to bias and design faults like finance, healthcare, and surveillance. These assessments should particularly assess the biases in data and outcomes, error margins and strict adherence with the prevailing data privacy and cyber security regulations. By proactive measures at this stage will help mitigate potential harm and foster

https://timesofindia.indiatimes.com/city/delhi/21-yr-old-arrested-for-blackmailing-college-student-with-ai-generated-explicit-photos/articleshow/117378489.cms

[28] Schwartz, O., *In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation*, IEEE Spectrum (17 August 2025), https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation

development of AI. Further, this will help in responsible and ethical use of AI technologies.

Conditional Legal Personhood to AI System

Considering the prolonged and complex nature of enacting a new AI legislation is an intricate task, a more feasible approach at this point would be to introduce the concept of "electronic legal agents" for highly autonomous AI systems. This could serve as a functional bridge. This concept would not fully equate AI with the identity of legal personhood but would bound AI systems ethically and contractually through their controllers and developers. This conditional recognition would ensure that AI induced crimes are not left unaddressed and someone remains accountable for its consequences.

AI Grievance Redressal Mechanism

India is on the edge of an AI revolution but the current legal infrastructure, specifically the Information Technology Act, 2000 falls short while addressing issues and challenges created by AI. There is a pressing need for the Central government to establish a dedicated central-level grievance redressal mechanism under the Ministry of Electronics and Information Technology. This body should allow the citizens to lodge complaints, and should also be empowered to investigate the cases and would facilitate fast legal proceedings wherever necessary. The existing IT Act's scope is not broad enough to address the unique nature of AI related offences. This would help in bridging the regulatory gaps and to uphold the spirit of justice.

Tiered Responsibility Model

With the advancement, AI systems are becoming more autonomous this raises a serious concern over past out-dated liability models. A tiered liability model is required to cater the issue of AI crimes with a balanced approach. The developers should be held accountable

and hence liable only when there is a flaw in product design or there exists a bias in the system. The developer's liability can arise when there is an issue with the implementation or the deployer lacked due diligence. Further, in case of intentional misuse the user should bear the responsibility. This layered classification would distribute the accountability as per the flaw and control level of the human agent involved.

CONCLUSION

The emergence of AI has raised a complex and evolving issue of how criminal liability can be imposed in cases where harm is caused by AI systems. With the increasing capabilities of technology, the existing criminal systems and legal framework remains far behind to keep up with the pace, particularly when AI systems lack consciousness, mens rea and the position of legal personhood. The traditional criminal law is primarily founded on the principle of act and intent of humans, but when applied to adaptive AI systems the traceability and foreseeability of action remains the biggest hurdle. Though there exist a few doctrines and models by which liability can be imposed on natural agents behind them but their application remains partial and differs as per context and circumstances. The question is not just how to impose liability, but how the issue related to AI crimes and harm is justly addressed. The harm caused by AI is increasing day by day be it deepfake scams, AI enabled cybercrimes or the autonomous vehicle collusions, which calls for a robust system that is more adaptive.

The recommendations proposed in the study aims to offer increments that can co-exist with current legal framework. The mandatory AI assessment prior to its deployment, granting electronic legal agent identity to AI, employing a tiered liability model and establishing a Central Grievance Redressal Mechanism will

represent a step towards making a responsive legal system.

These measures focus on constructing a futuristic framework that balances technological evolution with accountability and public trust. There is a need for gradual evolution and contextual adaptation of laws to meet the demand of technological revolution. In addition, interviews with the policymakers, law-enforcement agencies to uncover the ground level insights regarding accountability gaps and institutional preparedness. Only by imposing accountability, transparency and ethical checks at every level will we be able to build an ethical robust framework to address. Ultimately, the challenge does not lie in resisting technological change, but in ensuring that legal systems are too equipped to govern that justly and responsibly. Thus, as technology evolves and advances, the law must also evolve in parallel.