CIVIL AND CRIMINAL LIABILITY OF ARTIFICIAL INTELLIGENCE: RE-THINKING MENS REA AND LEGAL PERSONHOOD

Sivasankar S, SASTRA Deemed University, Thanjavur, Tamil Nadu, India.

Lakshmii Narasimhan S, SASTRA Deemed University, Thanjavur, Tamil Nadu, India.

ABSTRACT

The rapid proliferation of artificial intelligence (AI) in sectors ranging from autonomous transport to legal analytics has exposed deep gaps in the way civil and criminal law determine liability for AI-driven harm. This paper critically examines whether traditional legal frameworks centered on human agency, mens rea, and personhood—can meaningfully address incidents caused by AI, or whether new doctrines are required as autonomous AI systems capable of independent decision-making challenge these foundational principles.

Drawing on leading liability models, legal case studies, and comparative regulatory analysis, the paper explores how AI challenges core tenets of responsibility and accountability. It argues for a human-centric legal approach focused on clear oversight, prescribed negligence and strict liability standards, and mandated transparency, resisting the premature conferral of legal personhood on AI. The aim is to provide actionable legal perspectives that safeguard both innovation and public protection, advocating for adaptive frameworks capable of bridging the emerging "responsibility gap" in the age of autonomous machines. The paper further aims to address the evolution of ML & AI, its mechanism, applicability in business matters, courtroom practice in India and globally.

I. Introduction

The Artificial Intelligence though its application differs from each sector have given a self-explanatory statement as to its functioning. "It is received as AI is a large language model that uses deep learning techniques to generate human-like texts. It is based on the generative pre trained transformer (GPT) architecture, which uses a transformer neural network to process and generate texts. The model is pre-trained on a massive dataset of texts, like books, articles, and websites, so it can understand the patterns and structure of natural language. When given a prompt starting point, the model uses this pre-trained knowledge to generate text that continues the given input in a coherent and natural way."[1]

What is meant by "large learning language model?". It digests huge quantities of text data and infer relationships between words within the texts. The basic training given to the GPT is to predict the word in a sequence of words also known as next-token-prediction, which helps to predict the nest word in a sequence and masked-language-modelling, which helps to predict the middle word in a sequence[2]. Basically, what happens is the input data is processed individually and sequentially rather than as a whole corpus. This means that during training, the context window is fixed and only extends beyond a single input for a number of the phases in the process. This limits the complexity of the relationships between words and the meanings that can be derived.

In response to this issue, in 2017 a team at Google Brain introduced transformers[3]. Unlike LSTMs, transformers can process all input data simultaneously. The model can assign variable weights to various input data components in connection to any point of the language sequence by use of a self-attention mechanism[4]. This feature enabled massive improvements in infusing meaning into large language models and enables processing of significantly larger datasets. The first GPT was introduced in 2018 by an open ai named it as GPT1. and the model continued to evolve into a GPT-2 in 2019 and by 2020 with GPT model-3 and by 2022 it was evolved as

[1] Vaswani, A., et al., "Attention is All You Need," *Advances in Neural Information Processing Systems* 30 (2017): 5998-6008.
[2] Devlin, J., et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint* arXiv:1810.04805 (2018)
[3] Vaswani, *supra* note 1.
[4] Bahdanau, D., Cho, K., and Bengio, Y., "Neural Machine Translation by Jointly Learning to Align and Translate," *arXiv preprint* arXiv:1409.0473 (2014)

instructGPT and CHATGPT. [5]

Every single GPT model has a transformer architecture that is made up of an encoder to handle the input sequence and a decoder to construct the output sequence. The encoder and decoder both provide a mechanism for multi-head self-attention that enables the model to differentially weight various parts of the sequence in order to infer meaning and context. The encoder additionally uses masked-language modelling to comprehend the links between words and produce more intelligible replies. It is the self-attention mechanism that drives CHATGPT.[6] The following are the step-by-step process on how the AI generates the answers and suggestions,

1) It creates a query, key and a value vector for each token from the input it has received.

2) It calculates the similarities between the query from step one and the key of every other token by taking the product of two vectors together

3) Then it generates normalized weights by feeding the output from step 2.

4) It generates the final vector representing the importance of the token within the sequence by multiplying the weights (how much stressed the word was…) and finally give its suggestions.[7]

Article 3(1) of the EU AI Act, states 'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.[8]

II. Legal Personhood Concept in AI

Legal personhood is a fundamental concept that under Indian law determines whether an entity

[5] Brown, T., et al., "Language Models are Few-Shot Learners," *Advances in Neural Information Processing Systems* 33 (2020): 1877-1901; Ouyang, L., et al., "Training Language Models to Follow Instructions with Human Feedback," *arXiv preprint* arXiv:2203.02155 (2022).
[6] Radford, A., et al., "Improving Language Understanding by Generative Pre-Training," OpenAI (2018)
[7] *Id.*
[8] Regulation (EU) 2024/1689 of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (AI Act), arts. 3 (2024).

can have rights, obligations, ownership of property and also be liable. Two categories of persons are identified by Indian jurisprudence:

1. Natural persons – human beings;

2. Juristic persons – non-human entities that are recognized through legal fiction for the purpose of justice, convenience, and accountability (e.g., corporations, temples, idols, trusts, rivers).

With the advent of autonomous AI systems, the question arises if India will be able to acknowledge AI as a legal person by resorting to these doctrines.

A. Legal Fiction Doctrine (Indian Jurisprudence)

Legal fiction is one of the frequent sources Indian courts have recourse to, to bestow the non-human with personhood.

Non-exhaustive examples are:

- Corporations – simple & well-established juristic persons;
- Temples and Deities – acknowledged as legal persons (e.g. Yogendra Nath Naskar v. CIT, AIR 1969 SC 1089);
- Rivers and natural entities – briefly acknowledged as legal persons (e.g. Mohd. Salim v. State of Uttarakhand, 2017).

Presently, AI cannot be categorized as such since the objective of legal fiction in India is to bolster accountability. Granting AI personhood would in fact, diminish human accountability allowing developers or users to cloak themselves behind an "AI entity".

B. Separate Legal Identity Requirement

In India, the law insists that a legal person must possess the capacity to own property or assets (just like idol trusts and corporations), make contracts, initiate and be the target in legal proceedings and consequently, pay compensation.

On the other hand, AI cannot possess assets, make contracts on its own under the Indian Contract Act, 1872, or take on the loss financially. Accordingly, they do not pass the criterion for separate legal identity.

C. The Doctrine of Legal Fiction (Juristic Personality)

Non-human legal personhood is solely a product of legal fiction, where the law "pretends" to treat an entity as a human being for the sake of convenience, justice or accountability. This principle allows religious statues or companies to be entitled to rights and obligations even though they do not exist physically or biologically.

Currently, AI cannot be allocated personhood via legal fiction because this remedy would not be in line with the primary purpose of ensuring accountability or protecting social interest. Rather, it would cause the opposite effect by creating a scenario where the responsibility is shared among the manufacturer, the deployer, and the AI “entity”.

D. Functional Personhood / Instrumental Test

Some academics advocate for "functional" or "instrumental" personhood—imposing the limited personhood of AI if only it then is an instrument of accountability or risk management.

Through this approach, AI still remains a loser because:

- AI cannot assume any financial risk;
- The insurance or compensation process would still eventually involve humans or corporations;
- Granting personhood might lead to less accountability as it would allow human participants to escape liability.

Hence, functional personhood does not warrant the legal recognition of AI at this time.

AI cannot be considered for legal personhood at this stage under Indian legal doctrines — the legal fiction, attributability, separate legal identity, and functional accountability. The courts in India have only widened the scope of personhood to those entities which render the public good and enhance the accountability. The AI does not satisfy these criteria.

Hence, it is advisable for India to continue with human-centered liability and assign it to the developers, deployers, manufacturers, data trainers, and operators, rather than artificially creating an "AI person."

III. The AI Liability Challenge: Disrupting Traditional Frameworks

A. The Responsibility Gap:

Artificial Intelligence (AI) confuses the issues of causation by merging the human supervision and the algorithmic execution, which consequently makes the tracing of the decisions from which the harm emerges almost impossible. The "responsibility gap" signifies the situations where the exclusive blame cannot be directed at any human actor as the damage is caused by the self-governing and unpredictable behaviour of AI.[9]

Models of liability attribution:

- Strict product liability: Developers are responsible for design flaws.
- Fault-based liability: Deployers/platforms are accountable for not performing due diligence or for not taking measures against acknowledged risks.
- Shared liability: It proportions and distributes the accountability among the stakeholders based on their role and the risk they foresee/mitigate.[10]

Nowadays, advanced AI programs that utilize deep neural networks primarily work via the methodologies that resist straight forward causal analysis. Unlike giving constant rules, they take data as input to allow themselves to show their latent behaviour which cannot be connected to particular code, continually adapting during the process of deployment, and making their decisions at the scale where human supervision cannot keep up. This results in the "responsibility gap" the phenomenon in which harm is done but the conventional fault-based liability system cannot prove responsibility of human actors unambiguously.[11]

---

[9] Matthias, A., "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata," *Ethics and Information Technology* 6, no. 3 (2004): 175-183.
[10] European Commission, "Liability for Artificial Intelligence and Other Emerging Digital Technologies," Expert Group Report (2019).
[11] Bryson, J., "Patiency is not a Virtue: The Design of Intelligent Systems and Systems of Ethics," *Ethics and Information Technology* 20, no. 1 (2018): 15-26.

B. Doctrinal Challenges at the Core:

The complexity of causation: Artificial intelligence of modern times consists of several actors including data providers, algorithm developers, training engineers, deploying organizations, maintenance companies, and end-users where all of them contributes to the behaviour of the system. It becomes almost impossible to analyse and determine which contribution caused the harm proximately when it is compounded by AI's "black box" phenomena and data reliance.[12]

Foreseeability Dilemmas: The interaction of components in AI systems causes them to fail unexpectedly and to be attacked by unpredicted vulnerabilities. Besides, they can easily produce incorrect data through multi-system interactions and make it worse when data that is different from the training sets are encountered. The question that to be addressed: what level of AI unpredictability must developers reasonably anticipate?[13]

IV. Applying Traditional Civil Liability Doctrines

While Indian law has not yet dealt with AI-related precedent, it has nevertheless laid down principles which are applicable to the area of emerging technologies. The absolute liability doctrine that states “non-delegable and absolute” duty to ensure that the public does not suffer any harm. This also means that traditional defences can’t be invoked and compensation is to be paid in proportion to the capacity of the enterprise. This could be very much the case with AI, where faults leading to mass destruction of property and loss of life would eventually cause the companies applying such technology in critical areas like government, military, health, and transportation, to incur absolute liability. This is also in line with the standing of EU AI Act, which separates the AI in respect to their Harm or risk levels. This Absolute liability concept could be used when harm caused by AI is regarded not as a prohibited or high-risk category. The concept of strict liability may be applied in cases harm caused by AI which come under the Prohibited or High-risk category because this doctrine makes the person behind the AI liable without any exceptions.

The judgement in ***Puttaswamy v. Union of India (2017) 10 SCC 1***[14] wherein privacy rights were recognized also implied a duty on AI systems handling personal data. The enactment of

[12] Burrell, J., "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society* 3, no. 1 (2016).
[13] Amodei, D., et al., "Concrete Problems in AI Safety," *arXiv preprint* arXiv:1606.06565 (2016).
[14] *Justice K.S. Puttaswamy (Retd.) v. Union of India*, (2017) 10 SCC 1.

the Digital Personal Data Protection Act, 2023[15] now places obligations on "data fiduciaries" covering AI systems, with breaches attracting statutory and tort liability.

The path forward is to combine both:

- Adopt EU-style strict regulation for high-risk AI;
- Maintain US-style fault-based accountability for others

This dual approach observes innovation while ensuring legal responsibility.

C. Causation and Burden of Proof:

In the case of civil liability (tort and product liability) the claimant has to provide proof of:

1. Damage or harm,
2. Link between harm and the defendant's act or omission, and
3. Fault (in negligence) or defect (in product liability).[16]

But, A.I. has made things difficult in all three aspects:

- Opacity or black box effect: The victims are usually cut off from the algorithm or data that would have led to the decision being made.
- Autonomy: The human operator may not be aware of how the output was generated at all.
- Data dependency: The decisions are moulded by datasets and learning outcomes which are beyond the direct control of humans.[17]

So, it becomes very hard for the victims to establish fault or causation, particularly when AI acts in an unexpected manner.

[15] The Digital Personal Data Protection Act, 2023, No. 22 of 2023 (India).
[16] Fleming, J.G., *The Law of Torts*, 9th ed. (Sydney: Law Book Co., 1998), 193-225.
[17] Pasquale, F., *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge: Harvard University Press, 2015).

There is a proposal on reversal or soft-shifting the burden of proof based on economic efficiency and fairness.[18]

- Because AI operators or manufacturers have the upper hand when it comes to information and managing risk, they should be the ones to prove that no fault was committed.

- The victims need to present only the very least evidence of harm and a plausible causal link.

- The EU AI Liability Directive (2022)[19] which introduces:

"A rebuttable presumption of causality" when claimants are able to demonstrate non-compliance with safety or transparency obligations, is exactly the same as this approach.

In economic terms: shifting the burden results in the involvement of the parties that are most capable of preventing harm (developers, deployers) in taking precautionary measures.[20]

V. Case Studies

A. National Transportation Safety Board (NTSB) Highway Accident Report NTSB/HAR-19/03:

The 2018 Uber self-driving car accident in Tempe, Arizona, has been a turning point for the question of who is liable for the AI and the duty of care in the autonomous transportation sector. A self-driving Uber vehicle with a human safety driver in place struck and killed Elaine Herzberg on March 18, 2018[21]. The investigation uncovered a series of failures: Uber had turned off Volvo's collision avoidance system, the car's AI was not consistently recognizing Herzberg and did not foresee her crossing the road.

[18] Scherer, M.U., "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies," *Harvard Journal of Law & Technology* 29, no. 2 (2016): 353-400.
[19] Proposal for a Directive on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive), COM(2022) 496 final (European Commission, 2022).
[20] Calabresi, G., *The Costs of Accidents: A Legal and Economic Analysis* (New Haven: Yale University Press, 1970).
[21] National Transportation Safety Board, *Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian*, Highway Accident Report NTSB/HAR-19/03 (Washington, DC: NTSB, 2019).

The National Transportation Safety Board (NTSB) pointed out that Uber's lack of sufficient safety risk assessment and poor oversight protocols were the main factors behind the accident[22]. The safety driver was not paying attention, and the crash happened pretty suddenly she was watching television on her phone just seconds before the impact, attempted to intervene less than a second before the crash, and got no help from the AI system whose emergency braking was switched off and who was depending on the human to get involved last.

From a legal standpoint, the case set a clear precedent: companies cannot avoid liability by claiming that their systems were operating independently. It showed that the AI system's creators and users have to keep up the very high standards in testing, safety measures that are double or more, and human oversight that is strong and enforcing. The unfortunate incident has impacted future policy and regulatory talks about civil liability of self-driving cars and AI technologies worldwide.

B. The ongoing Tesla conflict:

The lawsuits that are still going on against Tesla regarding the "Autopilot" feature are trying to find out whether developers owe legal duties concerning the users' awareness of the system limitations[23]. The complainants say that Tesla failed to provide adequate care by advertising Autopilot in such a way that consumers would think the system would actually drive the car without any help from the driver when in reality the technology still required a very attentive and careful driver. The Jury verdict and settlements, including $243 million award[24], underscore that misleading advertisement and the failure to communicate the limitations of the system clearly can give rise to the setting up of unreasonable expectations and therefore to the developers being held liable for the harm caused even if the users are not clear about the features or they just rely too much on them. These cases are influencing the judiciary's way of handling AI-based product safety and accountability.

C. The German dispute on AI and Copyright:

Robert Kneschke v. LAION 2024[25] where the photographer Kneschke sued LAION, a non-profit organization that runs open datasets for AI research, where his image which is protected

[22] *Id.*
[23] *Banner v. Tesla, Inc.*, No. 20STCV48142 (Cal. Super. Ct. filed Dec. 18, 2020).
[24] Jury Verdict in *Molander v. Tesla, Inc.*, Riverside County Superior Court, Case No. RIC2005233 (Oct. 2023).
[25] *Robert Kneschke v. LAION e.V.*, Hamburg Regional Court, Case No. 310 O 227/23 (Sept. 2024).

by copyright, made its way to the LAION-5B database of LAION through web scraping that was automated. Kneschke tried to argue that his image was already in the database and that LAION had never got any rights or permissions to use it. LAION on the other hand contended that the Section 60d of the German Copyright Act was their shield, wherein that the nonprofit could copy the works for text and data mining in the name of scientific research which is also backed by the EU law.[26]

In September 2024, the Hamburg Regional Court decided that the actions of LAION were not a breach of the copyright, assuring that the agency's free and public interest datasets were ones that qualified for the research exception provided by German law[27]. The exception was limited to non-commercial and scientific purposes and LAION's being transparent and supporting research via reinvestment met these conditions. On the other hand, the court recognized that copyright holders could insist on their preferences through the use of machine readable opt-out methods including specific terms of use or technical measures like robots.txt files. The case denotes the setting of limits for AI dataset formation but at the same time leaves very important query regarding commercial use and future opt-out enforcements.

D. The Indian dispute on AI and Copyright:

The conflict between Asian News International (ANI) and OpenAI is based on the claims that OpenAI utilized ANI's intellectual property, which consists of news content protected by copyright, without the necessary permission to train its AI language model, the CHATGPT[28]. In the case initiated in late 2024, ANI alleged before the High Court of Delhi that OpenAI has infringed its copyright by improper use of its content including the subscription-only content and that it has inflicted economic damage to ANI by the unauthorized use of its intellectual property. ANI is asking for an injunction to stop OpenAI from using its news and seeking monetary compensation for the claimed infringement.

OpenAI's rejected the accusations of copyright infringement, arguing that its model performs analysis on public data and thus, is not copy in the sense of reproducible diffusion of exact articles. The company maintained that using material that is freely accessible online (publicly

[26] German Copyright Act (Urheberrechtsgesetz - UrhG), § 60d; Directive (EU) 2019/790 of the European Parliament and of the Council on Copyright and Related Rights in the Digital Single Market, art. 3-4.
[27] *Robert Kneschke*, *supra* note 28.
[28] *Asian News International v. OpenAI OpCo LLC & Ors.*, CS(COMM) 687/2024 (Delhi High Court, filed Nov. 2024).

available documents) is fair use and also questioned the competence of Indian courts on the matter, asserting that the servers and main offices are outside India. [29]

The case attracts the participation of amicus curiae with whom the courts have been in contact for the technical studies on the copyright exceptions appliable to AI training and the jurisdiction of courts. Moreover, the industry groups like the Federation of Indian Publishers and the Digital News Publishers Association have become part of the process, thus emphasizing the far-reaching consequences the case would have not only for AI but also for copyright law and the media sector.[30]

The Delhi High Court currently considering the question of whether OpenAI's AI training methods infringe copyright or are fair use, thus making this case a landmark for copyright and AI liability in India.

The matter brings into focus the problem of international disputes with respect to jurisdiction determined by the location of the injury and the defence contention that the AI's independent action does not trigger corporate liability which is already affected by various unrelated factors including those coming from the end users.

VI. The EU Draft Act on Liability

A. The AI Act (Regulation 2024/1689): Risk-Based Classification:

The AI Act is yet to come and introduces the notion of risk into tiers as Prohibited (unacceptable risk like social scoring), high risk (subject to strict obligations), limited risk (just transparency requirements), and minimal risk (no obligations). [31]

One of the High-Risk AI Obligations is the imposition of high-end risk management systems, data governance quality controls to ensure the elimination of representative and biased datasets, technical documentation support, automatic event logging, transparency about capabilities and limitations, human oversight design, and accuracy, robustness and cybersecurity standards[32]. It

[29] *Id.*
[30] Federation of Indian Publishers and Digital News Publishers Association, Submissions as Interested Parties in *ANI v. OpenAI* (Dec. 2024).
[31] Regulation (EU) 2024/1689 of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (AI Act), arts. 5-7, OJ L 2024/1689 (2024).
[32] *Id.*, arts. 8-15.

also includes the post-market requirements such as quality management, monitoring, incident reporting, and taking corrective action.

Liability Implications: Apart from the regulation aspect (fines can be as much as €35 million or 7% of global turnover), compliance demarcates the civil liability standards. The Act's stipulations are the least of the care standards, whereby infractions might be considered negligence per se. The necessary documentation is a source of causation proof, while the harmonized standards minimize dubiety.[33]

B. The EU AI Liability Directive (AILD):

The AILD proposes to deal with one of the major problems of black-box and other AI systems in general including proof difficulties, at the same time by creating a collateral mechanism which can be used in these cases: a rebuttal presumption of causality.[34]

The presumption of causation stated in the AILD is between the fault of the accused (for instance, developer or owner) and the harm done if the following conditions are fulfilled all of them:

1. The plaintiff must show that the accused did not meet his duty of care (or the fault is otherwise presumed).

2. The fault is considered to have had a power of influence that was reasonably likely to sway the AI system's output.

3. The claimant proves that the AI system's output or the failure to produce an output was the cause of the damage.[35]

Such a mechanism does not just reverse the burden of proof but it is also a robust turning of the tables in Favor of the claimant in risky situations. The law here compels the developer to show through internal records, etc., that they were not at fault in respect of the high-risk AI system, which ultimately leads to the rebuttal of the presumption of fault. Such a systematic

[33] *Id.*, art. 99.
[34] Proposal for AI Liability Directive, *supra* note 22, art. 4.
[35] *Id.*

pressure brings along the requirement for transparency, which is not achieved through negligence law.

C. The Complementary Role of Product Liability and No-Fault Schemes:

The AILD is supported by the new Product Liability Directive (PLD), which gives the right to the individuals for claiming compensation for damage caused by a defect on the basis of the strict liability.[36] Nevertheless, a thorough strict liability system might not be able to reimburse, in cases where the AI makes a mistake but the manufacturer and the developer observed all the scientifically recognized standards and were even more careful than required. The classical concern to identify a "debtor" is not enough when the damage is absolutely not due to human error.

In such cases of residual losses where no negligence, imprudence, or unskillfulness can be proven, a regulatory evolution is needed, which will be a shift from traditional civil liability to one of financial management of losses. This aids the establishment of No-Fault Redress Schemes, which take on the role of social insurance by ensuring that victims get compensation irrespective of proof of fault[37]. This duality of strict liability for traceable defects or errors and no-fault schemes for genuine autonomous accidents, maintains compensatory fairness throughout the entire spectrum of algorithmic harm.

VII. Conceptual Tensions

In recent times, Artificial Intelligence (AI) has been an important factor in contributing to major technological developments on a global level and in transforming / revolutionising the field of cyber security and cyber space. On a side note, AI systems have also made autonomous decisions thereby causing harm to people. Attributing Civil and Criminal liability to AI driven models and systems creates conceptual conflicts with current legal systems, which require a review of basic principles, including *mens rea, actus reus* and legal personhood.

A. *Mens rea* and *Actus reus* in autonomous systems:

It is quite easy to link an AI's actions to *actus reus*, the physical act or omission that constitutes

[36] Directive (EU) 2024/2853 on Liability for Defective Products (Product Liability Directive), OJ L 2024/2853 (2024).
[37] Wagner, G., "Robot Liability," in *Liability for Artificial Intelligence and the Internet of Things*, ed. S. Lohsse, R. Schulze, and D. Staudenmayer (Baden-Baden: Nomos, 2019), 27-62.

a crime (e.g., an AI-driven vehicle causes a collision). The primary conflict is around *mens rea*—the "guilty mind" or criminal intent[38]. Traditional legal doctrines require a purposeful and blameworthy state of mind (intent, knowledge, recklessness). AI systems grounded in algorithms lack consciousness and moral judgement and, as such, do not create *mens rea* in the traditional sense.

For example, in the UK case of ***R v. Deputy Governor of Parkhurst Prison***[39], the court stressed the human element of intent in criminal liability. The concept of AI liability remains a grey area due to the absence of human-like intent when the theory as stated in the above case is linked to an AI system/model. The AI acts as an "innocent agent," performing the *actus reus* (the act) without *mens rea* (the intent), complicating direct criminal attribution.

An AI system can create the *actus reus* (for example, an autonomous vehicle which is the driver in a fatal collision like the Uber self-driving car in Arizona in 2018)[40], but the way the case is viewed legally is with respect to *mens rea*. AI systems make determinations based on complicated algorithms, data processing, and machine learning; however, AI systems do not possess a conscience or moral sense or human-like capability to intend or possess a "guilty mind". Even if their behavior appears "autonomous," everything they do is a calculation or computational result of an event, not a self-motivated act. AI's "black box" problem, where the decision-making process of an AI is, in many cases, unknown even to its developer, complicates locating the requisite mental state to a human actor in the eyes of the law[41].

The concept of *mens rea* covers a variety of mental states. Intention, knowledge, recklessness, and negligence are the levels of guilt. The Bharatiya Nyaya Sanhita, 2023, highlights these distinctions using terminology such as "voluntarily," "intentionally," and "knowingly." ***Section 33*** of Bharatiya Nyaya Sanhita[42] defines voluntariness as causing an effect by means intended or known to be likely. This formulation encompasses the essence of conscious moral choice.

The *actus reus* element must be a voluntary act and mere bodily movement does not constitute an offence when the act / omission is involuntary in nature. This is based on the premise that

[38] International Journal of Research Publication and Reviews, Vol 5, no 11, pp 1886-1891 November 2024 – Criminal Liability of Artificial Intelligence by Teena Arora and Dr. Shailja Thakur.
[39] R v. Deputy Governor of Parkhurst Prison [1991] BCC 713.
[40] The Uber Autonomous Car Accident – Naavi.org
[41] ILI Law review, Summer Issue 2020 - Artificial Intelligence: The Liability Paradox by Gyandeep Chaudary.
[42] Section 33 of Bharatiya Nyaya Sanhita, 2023.

"a guilty mind must actuate the guilty act". Therefore the *mens rea – actus reus* framework focuses on the conscious choices in turn deterring the offenders according to the level of culpability.

B. Can AI form intent? Comparison to Corporate Liability under IPC and BNS:

AI cannot embody emotional intent in the human psychological sense. To address the issue, legal scholars often look into the corporate criminal liability model, as in the liability of other non-human legal business entities.

Indian Corporate Liability Model: In Indian cases, including the landmark cases of ***Iridium India Telecom Ltd. v. Motorola Inc.***[43] and ***Sunil Bharti Mittal v. Central Bureau of Investigation***[44], the Supreme Court explained that corporations can be criminally liable for offences that require *mens rea*. This is accomplished through the "doctrine of attribution," or "alter ego," which means the corporation itself will be implicated in the acts and intentions of key individuals of the corporation (directors, senior management, etc.)[45]. In the ***Standard Chartered Bank v. Directorate of Enforcement***[46] case, the application of this doctrine was further clarified, and also, it was specified that a corporation can be fined for criminal offences where imprisonment is mandated.

The Indian Penal Code (IPC) and the Bharatiya Nyaya Sanhita (BNS), which was enacted recently repealing IPC, are primarily geared towards attributing liability to human activities and do not provide specific provisions for AI offences. Although the agency of the corporate model may be a useful starting point, it is highly problematic in application to AI models as companies are recognised as an "artificial person" under **section 2(20)** of the companies act and the same concept cannot be attributed to AI systems or models.

- Attribution Gap: The alter ego doctrine focuses on the identification of a human "directing mind." When it comes to true AI independence, where the AI makes a completely unpredictable decision, that is the result of machine learning after deployment (for example, the Knight Capital Group's AI trading algorithm lost $440

[43] Iridium India Telecom Ltd. v. Motorola Inc (2011) 1 SCC 74.
[44] Sunil Bharti Mittal v. Central Bureau of Investigation (2015) 4 SCC 609
[45] International Journal of Research Publication and Reviews, Vol 5, no 11, pp 1886-1891 November 2024 – Criminal Liability of Artificial Intelligence by Teena Arora and Dr. Shailja Thakur.
[46] Standard Chartered Bank v. Directorate of Enforcement (2006) 4 SCC 278.

million because of a bug in the software), it is nearly impossible to identify a blameworthy human as a decision-maker.

- Perpetrator via another: Most AI behaviour is still classified as "perpetration via another," which means that the AI is simply an innocent agent and the human programmer/user is the violator, especially when the outcome was foreseeable through negligent or defective design.[47]

Section 3(5) of Bharatiya Nyaya Sanhita[48] extends the application of the act to offenses committed by "any person in any place without and beyond India committing offence targeting a computer resource located in India." Therefore, if an AI system physically situated overseas damages Indian computer resources, it could be held liable. As a result, cross-border AI activities may have extraterritorial applications. Though, mechanisms for enforcing such provisions still remain inadequate.

VIII. Corporate Criminal Liability and the Concept of Vicarious Liability – The Analogy

The application of corporate criminal liability in India provides a vital analogy for addressing the accountability gaps surrounding AI. The Indian judicial system has expanded its application of attribution doctrine through courts to extend *mens rea* (guilty mind) and vicarious liability principles to artificial legal persons.

In ***Iridium India Telecom Ltd. v. Motorola (2010)***[49], the Supreme Court delivered a landmark judgment that solidified a corporation's criminal liability, even for offences requiring *mens rea*. The court established two legal principles which base corporate accountability on the criminal conduct of directors and senior managers who control company decisions. This decision, stemming from charges of cheating and criminal conspiracy, was pivotal in allowing corporations to be prosecuted for a full range of offences under the Indian Penal Code, including those with mandatory imprisonment provisions.[50] AI systems are frequently used within corporate entities as products or services. The Sanhita corporate liability paradigm may ascribe AI-related injuries to corporate entities. This may bypass the issue of determining an

[47] The Amicus Qriae - Artificial Intelligence and the Death of Mens Rea: A Legal Dilemma by Shivangi Kumari.
[48] Section 3(5) of BNS, 2023
[49] Iridium India Telecom Ltd. v. Motorola Inc (2011) 1 SCC 74.
[50] LawFoyer International Journal of Doctrinal Legal Research, Volume 3 Issue 1 - Criminal Accountability For AI: Mens Rea, Actus Reus, And The Challenges Of Autonomous Systems by Akanksha Priya.

AI system's direct mens rea. Corporate knowledge or intent may substitute for algorithmic mental states when the above case is applied to an offence caused by an AI system.

The Bombay High Court's earlier ruling in ***State of Maharashtra v. Syndicate Transport Company Pvt. Ltd.***[51] laid essential groundwork for this principle. The court examined a shareholder who committed fraud while working for the for the entity in this case. The court decided that while corporations can be held liable for certain offences, they are exempted from liability when such offences necessitate involvement of a natural person or when those offences are punishable with imprisonment.

The cases together demonstrate how Indian law has developed through various stages of legal development. These cases show how the judiciary overcomes the conceptual barrier of non-human entity lacking "mind" by attributing the mental element of human intent and actions to legal persons. The mechanism provides a useful way to understand AI liability because it links AI actions to human programmers and deployers and corporations which allows current legal systems to handle AI-related harm.

IX. Current Legal Framework in India

Apart from Bharatiya Nyaya Sanhita, there are various statutes in India governing offences such as Information Technology Act, 2000 and Digital Personal Data Protection Act, 2023 covering cyber crimes and data protection standards respectively.

- **Information Technology Act, 2000** –

    - **Section 66**[52] imposes criminal liability for unauthorised access or damage of computer resources. This provision is applied only when the outcome of the intrusion is intended and purposeful and therefore fail to addresses cases where the harm or damage maybe caused by an AI system that is usually unintended in nature.

    - **Section 43A**[53] imposes liability on Body Corporates for failure to maintain reasonable security practices. The provision is narrow in nature converging on

[51] State of Maharashtra v. Syndicate Transport Company Pvt. Ltd. AIR 1964 Bom 195.
[52] Section 66 of Information Technology Act, 2000
[53] Section 43A of Information Technology Act, 2000

sensitive data protection and does not address broader algorithmic decision harm.

- **Section 43**[54] Penalizes unauthorised access and data theft using computer resources. AI systems are autonomous decision making systems and are therefore capable of producing unpredictable outcomes including data theft or unauthorised access to resources that require subscription or the relevant key.

- **The Reasonable Security Practices Rules, 2011**[55] under the IT act provides further guidance by establishing data protection obligations under the IT Act. However the focus is only on information and data security rather than algorithmic decision making including its transparency or fairness.

The IT Act was drafted before modern AI era and does not adequately address its complex characteristics, such as self-learning and autonomous features, even though AI typically fits the current definition. This raises questions about things like intellectual property rights for content created by AI, culpability for damage produced by autonomous AI, and the necessity of particular regulations on algorithmic accountability.

- **Digital Personal Data Protection Act, 2023 –**

The act represents India's first comprehensive data protection regime establishing principles such as purpose limitation, data minimization and quality requirements. It governs and imposes obligations on entities processing personal data.

The DPDP Act indirectly oversees data-fed AI systems, which must process personal data in a lawful and protective manner. However, the law governs input data handling rather than algorithmic outcomes and therefore, it cannot deal with harms that result from biased or wrongful AI decisions.

The **Ministry of Electronics and Information Technology** released **"National Strategy for Artificial Intelligence"** in 2018[56]. In 2021, the **NITI Aayog** published the

[54] Section 43 of Information Technology Act, 2000
[55] Reasonable Security Practices Rules, 2011 under Information Technology Act, 2000
[56] National Strategy for Artificial Intelligence (#AIFORALL) by NITI Aayog

'Approach Document for India Part 1'[57]. These policy documents provide governance visions without defining enforceable norms. They acknowledge ethical considerations but lack legal power. This presents an enforcement gap in Indian AI governance.

As per **Section 2(b)**[58] of the act, the term "automated" is defined as "means any digital process capable of operating automatically". Further, **Section 2(x)**[59] defines the term "processing" to include "wholly or partially automated operation or set of operations" Therefore these definitions are inclusive in nature to include the acts of AI systems.

X. Foreign Jurisprudence and Theoretical Models – A comparative analysis with the Indian Perspective

The fast development of artificial intelligence demands a comparative study of legal systems which analyse how various countries handle AI criminal responsibility through the examination of foreign judicial decisions and academic models against Indian human-based legal frameworks.

A comparative study of foreign and Indian legal regimes on AI and criminal liability highlights significant theoretical and practical differences, which reflect a wider theme of legal and policy discussions about assigning culpability for acts executed by machines effectively, acting autonomously. The traditional criminal law framework in India, often based upon *actus reus* (physical act) and *mens rea* (guilty mind), struggles with the idea of "machine guilt" because machines cannot have intent, morality, or consciousness. Around the world, several countries have at least begun to create various new risk-based regulatory schemes, but no broad international standards appear on the horizon.[60]

- **The European Union's Legal Framework:**

  The EU is at the forefront of comprehensive regulation, with the **European Union AI Act** being the most high-profile example. The Act came into force in the year 2024 and is concerned with AI regulation from a risk-based perspective. The Act regulates AI usage and market access, while the corresponding AI Liability

[57] NITI Aayog, RESPONSIBLE AI #AIFORALL: APPROACH DOCUMENT FOR INDIA PART 1, 12-18 (2021).
[58] Section 2(b) of Digital Personal Data Protection Act, 2023
[59] Section 2(x) of Digital Personal Data Protection Act, 2023
[60] Vintage Legal – Criminal Liability of AI in India by Yuwaraj Yadav.

Directives propose to ease the burden of proof for victims with a "presumption of causality" in civil cases. For criminal liability, the EU's current approach primarily delegates to national laws and the doctrine of attribution, delegating liability back to humans and/or corporate entities. Furthermore, there are ongoing debates regarding "electronic personhood," but only in a civil context and only for AI "high risk."

The EU has put into place an extensive, forward-thinking strategy, with its flagship, the EU AI Act. This Act classifies AI systems based on risk (unacceptable, high, limited, minimal) and creates obligations for high-risk systems, such as those used in critical infrastructure or law enforcement. Although the Act is primarily concerned with regulatory compliance and safety within the marketplace, it is accompanied by draft civil liability directives that create a presumption of casuality for harms caused by AI, limiting compensation claims to negligence absent proof of intention. This strategy prioritizes safety and human rights within a top-down framework of regulation.[61]

- **The United States Legal Framework:**

U.S. law primarily relies on existing common law principles, focusing on product liability, negligence, and tort law. Criminal liability for AI incidents often defaults to the human operator or manufacturer based on the foreseeability of risk and duty of care. Case law such as the **Uber autonomous vehicle incident (2018)** highlighted gaps, but criminal prosecution focused on the human safety driver's negligence, not the AI. There is no current federal AI-specific criminal statute.

The US model is less centralized, relying on a hodgepodge of state laws and common law doctrines (in tort law, product liability, negligence etc). There is no federal AI law. The legal cases involving harm caused by AI (i.e. the Uber self-driving fatality in Arizona in 2018) are primarily concerned with human negligence or corporate liability. The focus remains on whether the human programmer, manufacturer or operator acted reasonably under the standard of care owed to the

[61] EU AI Act: first regulation on artificial intelligence – Topics (European Parliament).

injured party. Liability is often determined under cases of product defect (for instance, the California strict liability rule in the case ***Greenman v. Yuba Power Products, Inc.***[62] set the precedent for strict no-fault product liability in California [this was not an AI case, but established the principle]), therefore avoiding the cumbersome question of intent by the AI. [63]

Indian law, including the Bharatiya Nyaya Sanhita, 2023 (previously referred to as the Indian Penal Code) is based on human characteristics of intent and conduct as the factual basis for criminal liability. The statute primarily assumes human agency primarily focusing on the aspects of *mens rea* and *actus reus*. AI systems or entities, that basically lack the criminal intent i.e., the basic ability to think and act on its own, can therefore not be held liable as "persons" in any manner under the Indian legal perspective[64]. In the Indian legal landscape, it is the corporation that may be held vicariously liable for AI-caused harm and therefore the AI itself is not considered to be a legal person for criminal purposes. The liability typically falls on the developers, creators and the users for AI driven harm under the doctrines of vicarious liability and negligence. No Indian courts have acknowledged criminal liability on AI itself; the legal debates remain largely theoretical. Notably, there are no precedents related to AI-specific liability that have been rendered or passed by the courts in India.[65]

Currently, the Indian Criminal Laws are insufficient in addressing harms caused by AI systems as AI is not a person and further lacks moral conscience and intent. There is a need for further guidelines to be framed by the government particularly concerning harms caused by AI and the attribution of liability for specific offences concerning various domains. The EU has formulated a risk based regulatory framework that classifies risk into 4 tiers. Based on these tires, the conformity assessments, human oversight and transparency requirements differ. The US has domain specific sectoral regulations that govern AI caused risk. Similarly, India may formulate a framework/guidelines integrating the risk based model of EU and the sector/domain specific regulation of US. Impact assessments, certification requirements and industry standards could prevent harms proactively.

[62] Greenman v. Yuba Power Products, Inc. 59 Cal. 2d 57
[63] The Amicus Qriae - An analysis of the Impact of Artificial Intelligence on legal liability: A comparative study of Indian and International jurisprudence by Sharmila Solanki
[64] Lawfullegal - Can Machines Be Guilty? Reimagining Liability in the Age of AI by Deepak Kumar Gupta
[65] Nirma University Law Journal: Volume-8, Issue-2, July-2019 - Criminal Liability Of The Artificial Intelligence Entities by Ankit Kumar Padhy & Amit Kumar Padhy.

Conclusion:

To sum up, the quick development of Artificial Intelligence puts the basic ideas of civil and criminal liability most notably notions of *mens rea* and legal personhood under a serious challenge. Existing laws are still far from being fully equipped to deal with the problem of AI's independent decision-making and the resulting "responsibility gap". Hence, it is an inevitable arrangement of strict and vicarious liability, disclosure requirements, and human supervision that will keep the AI accountable while allowing the AI to be used in a socially responsible manner.