THE USE OF AI IN WHITE COLLAR CRIME: WHO'S LIABLE WHEN MACHINES MISBEHAVE?

Ridhikamini Basu Mallick, Jindal Global Law School (JGLS)

ABSTRACT

The integration of Artificial Intelligence (AI) into corporate environment has not only transformed how business operate but has also redefined an entire web of white-collar crimes. While AI promises predictive high-tech foreign investment, insights, efficiency and support, it also brings in algorithmic manipulation, breach of privacy, infringement of rights and regulatory misconduct.

The central aim of this paper is to deliberate how AI tools are being leveraged, intentionally or inadvertently, in facilitating white collar offences. Can a machine "intent" to commit a crime by itself? Who holds culpability when an algorithm makes a harmful and illegal decision? Where do we draw the line between the system error and criminal liability?

This paper shall try to assess how traditional frameworks of criminal law are challenged by AI's role and will brief examine whether the existing statutes are well equipped to respond to AI-enabled offences. This paper seeks to give a brief idea into the criminal implications of artificial intelligence in corporate world and shall hope to call upon a more nuanced and effective legal framework to tackle this new-age technology.

Keywords: Artificial Intelligence, White Collar Crimes, Digital Age, Surveillance, Status quo

INTRODUCTION

The anatomy of corporate crime has undergone a vast change in this constant evolution of global capitalism and technological innovation. White collar crime, which was once understood as acts of fraud or embezzlement, now emerges from the very design of corporate infrastructure, business practices and digital platforms. Parallelly, the integration of artificial intelligence (AI) into business management and surveillance systems has introduced a new area of study in field of criminology. Within a decade, an increasing number of AI related incidents have emerged, demonstrating their exploitations of norms and harm to the common people, including discriminatory and dissented breach of individuals' privacy, data, employment, housing and medical care.[1]

Which leads us to think; in an era where machines increasingly facilitate both control and error, how do we ensure that innovations don't come at the cost of justice and security? who is to be held responsible for a crime orchestrated or masked by a machine? And how does one trace intent, causality and culpability in systems designed to minimize human oversight?

This paper shall explore the answer to these questions and raise some, whilst arguing that contemporary corporate crimes cannot be understood outside the algorithmic systems that increasingly govern and assist in decision-making, compliance and accountability. The main proposition of this paper is that artificial intelligence, as we know it, is not merely a tool for compliance or a machine to assist humans in their day-to-day events, but can also become complicit in, or even central, to the acts of fraud, manipulation of data and concealment of criminal intent. Thus, the stakes are high and real.

2. Artificial Intelligence (AI) and White-Collar Crime (WCC)

When AI recommends harmful content or performs discriminatory actions, it reflects not just the will of the machine, but the choices and institutional goals set by humans, often rooted in profit motives or operational efficiency[2]. Culpability, then, becomes an inquiry into not just *who* did something wrong, but *how* systems were designed to deflect responsibility. The

[1] Karl Manheim & Lyric Kaplan, *Artificial Intelligence: Risks to privacy and Democracy*, 21 YALE J.L & TECH. 106 (2019)
[2] Shoshana Zuboff, *The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power* (New York: PublicAffairs, 2019)

"*problem of many hands*"[3] plagues automated environment where multiple actors, from data annotators and engineers to executives and vendors, contribute to the outcomes, none of whom claim full responsibility.

Since the paper deals with the presence of AI in corporate sectors, it cannot ignore that corporate stakeholders have also become victims to algorithmic opacity in business. It would also be careless to think that only AI, in this regard, is to be blamed. AI is a product of human mind that was given birth to, with the aim to enhance efficiency, reduce human error and optimise resource allocation. It is the programmers who write the initial code, engineers who build the data, and corporate stakeholders who determine how and where the AI systems are implemented. It can be said that AI is an extension of human intent shaped by the values, biases and limitation of those who create and control it. So, when AI make decisions, particularly in sensitive areas like financial compliance or surveillance, they often reflect implicit assumptions and priorities encoded into them by their developers. Further, AI has become a core instrument in corporate surveillance, running thousands of coding to maintain and tackle fraud detection, hacking and network breaches. Firms nowadays increasingly rely on AI as a weapon to gain advantage over competitors, which is producing several deleterious outcomes for corporate stakeholders, the capital market and the employment rate of the country[4]. These criminogenic corporate structures are not merely passive settings in which WCCs occur but active contributors to criminal activities.

A report by the World Economic Forum Global Risks Report 2025[5] comments that AI *"has the potential to blur boundaries between technology and humanity"* and that misinformation and disinformation is magnified by the widespread adoption of generative AI to produce what is known as *"Synthetic content",* which ranges from deepfake videos, voice cloning, the production of counterfeit websites, weaponization of AI in military, and criminal use of AI to initiate cyberattacks.

It is also important to mention how AI is being rapidly used as customer service agents to hyper realistic character chats. What began as utility-driven tools has quickly expanded into emotionally, responsive, personalised and even addictive forms of engagement. Companies

[3] Nissenbaum, H, 1996, 'Accountability in a Computerized Society', *Science and Engineering Ethics,* vol. 2, no. 1, pp – 25-42. https://doi.org/10.1007/BF02639315
[4] Katia Porzecanski, *JPMorgan Commits Hedge Fund to AI in Technology Arms Race,* BLOOMBERG (2019)
[5] Marl Alsner, Grace Atkinson, Saadia Zahidi, *Global Risks Report 2025*, World Economic Forum, January 15, 2025.

now deploy AI personas that mimic empathy, save long-term memory and adapt tonally to users, often without clearly disclosing that the interaction is entirely artificial, not just to serve but also to retain users' biodata. Deepfake technology, *"a new tool for an old age problem of disinformation"*[6], which may not seem as harmful as other white-collar crimes but has created a new level of threat to the autonomy and privacy of the common folks. Deepfake was used originally by the movie industry for entertainment purposes, however recent events showed that it has now become a tool to tarnish someone's reputation and defraud companies. A multinational firm based in Hong Kong lost $25 million after an employee was duped by a deepfake video conference featuring avatars of real colleague[7]. According to McAfee survey[8], more than 75% of Indians have reported seeing deepfake content, 22% encountered political deepfake and 38% have been victim to deepfake scam. Many assumed a celebrity deepfake for the real thing. One can recall a deepfake video of actress Rasmika Mandhana circulating on the internet[9]; the video had her face wrongly put on another person's body. Furthermore, deepfake can also tamper with evidence in courts to get favourable result. In UK, a deepfake audio was presented as evidence by a mother in a child custody case, where the father was shown as an abusive person to support her claim for custody[10].

From a different view of point, this also poses two-fold harm: first, the psychological exploitation, as users are nudged toward certain behaviours or beliefs under the guise of companionship; and second, economic exploitation. Such AI technology have also been used to tamper with the fair and free elections, by misleading the people and influencing their mindset[11] and posed a corporate threat by deceiving a bank executive in the UAE using a false

[6] James Andrew Lewis, *"Trust your eyes? Deepfakes Policy Brief"*, Centre for Strategic & International Studies. Washington DC, October 23, 2019.
[7] Karen Cheung, '*Hong Kong Employee Duped into Paying $25 Million in Deepfake Scam*', Hong Kong Free Press (February 5, 2024) https://hongkongfp.com/2024/02/05/multinational-loses-hk200-million-to-deepfake-video-conference-scam-hong-kong-police-say/
[8] McAfee Corp., McAfee Survey 2024, April 18, 2024, Press Release. https://www.mcafee.com/de-de/consumer-corporate/newsroom/press-releases/press-release.html?news_id=4698979d-2a55-4f71-84be-c04b41fc7bdc
[9] The Hindu, *"Delhi Police Arrest Techie From Andhra Pradesh For Rashmika Mandanna Deepfake Video"*, January 21, 2024. https://www.thehindu.com/news/cities/Delhi/delhi-police-arrest-techie-from-andhra-pradesh-for-rashmika-mandanna-deepfake-video/article67760419.ece
[10] Gabriella Swerling, *"Doctored Audio Evidence Used To Damn Father In Custody Battle"* (2020) The Telegraph, https://www.telegraph.co.uk/news/2020/01/31/deepfake-audio-used-custody-battle-lawyer-reveals-doctored-evidence/
[11] Al Jazeera, Yasraj Sharma, "*Deepfake Democracy: Behind the AI trickery India's 2024 election"* (2024) https://www.aljazeera.com/news/2024/2/20/deepfake-democracy-behind-the-ai-trickery-shaping-indias-2024-elections

voice and taking $35 million[12].

In addition to the malevolent acts, gendered exploitation is rapidly evolving with technology. It is no longer necessary to physically capture images of women; one can now fabricate erotic content entirely through AI, turning revenge profit and objectification into low-cost crimes.

A very recent case comes to mind with respect to this[13]. A viral Instagram account called "Babydoll Archi" amassed over 1.4 million followers by posting seductive reels and curated imagery branded as Archita Phukan. The account as launched in August 2020 but gained traction only in late June-July 2025, after posting a viral reel synced to "Dame Un Grrr", followed by AI-generated content featuring adult star Kendra Lust. Following a cyber defamation complaint filed by the woman whose image was used, investigations revealed that Archita Phukan never existed. The persona was entirely fabricated using AI tools like Midjournery AI, Desire AI, OpenArt AI and OpenAI, based on a single image of the complainant from Dibrugarh, Assam. The perpetrator, Pratim Bora, a mechanical engineer and former ex-partner of the complainant, created the account initially as digital revenge. It later became a profitable venture, yielding up to ₹10 Lakh via adult-content reel views and pay-to-access links like "Actual fans" (a mimic of OnlyFans) on Linktree[14]. An FIR was registered under relevant sections of the BNS, 2023. This incident sparked nation-wide debate on the ethical boundaries of generative AI and why Indian laws lack account for AI-generated imagery and deepfake pornography.

So, what legal apparatus India has to offer?

The replacement of the colonial-era three criminal laws with the new ones namely, Bharatiya Nyaya Sanhita 2023 (BNS), Bharatiya Nagarik Suraksha Sanhita, 2023 (BNSS) and Bharatiya Sakshya Adhiniyam 2023 (BSA) introduces new statutory emphases that may affect the way white collar crime is prosecuted.

---

[12] Thomas Brewster, "*Fraudsters Cloned Company Directors Voice In $35 Million Heist, Police Find"*, (2021) https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/

[13] India Times, Nancy Jaiswal, *"Babydoll Archi Wasn't Real: How A Viral AI Instagram Star Was Built on One Real Woman's Photo for Fame, Revenge and Profit*", July 15, 2025. https://www.indiatimes.com/trending/babydoll-archi-wasnt-real-how-a-viral-ai-instagram-star-was-built-on-one-real-womans-photo-for-fame-revenge-and-profit-663942.html

[14] India Today, Priyanjali Narayan, *"Ex-Lover Turns Revenge into Porn Profit, Morphs Assam Girl Into Babydoll Archi"*, July 14, 2025. https://www.indiatoday.in/india/story/lover-revenge-plot-turned-profit-babydoll-archi-adult-entertainment-star-kendra-lust-morphed-ai-images-assam-girl-2755445-2025-07-14

While the definitions of cyber-fraud, impersonation and document tampering have been expanded, they still rely on traditional legal concepts being retrofitted to digital realities. Given the global trends, such as the EU AI Act, OECD's AI principles and the US Algorithmic Accountability Act, India could have used this opportunity to introduce even a basic monitoring statute on AI systems used by the corporations and commoners; yet it remains fundamentally anthropocentric and reactive. This is reflected in the Ministry of Electronics and Information Technology (MeitY)'s response to a question in LokSabha in 2023 that "various central and state government departments and agencies have commenced efforts to standardize responsible AI development, use and promote the adoption of best practices". However, the Ministry added, "the government is not considering bringing a law or regulating the growth of artificial intelligence in the country"[15].

On the other hand, NITI Aayog's National Strategy for Artificial Intelligence in 2018 and the "Principles for Responsible AI" in 2021 outlined the ethical standards focusing on accountability, privacy, transparency and security in AI based applications[16]. India also made regulatory efforts, including the proposed Digital India Act and the establishment of the Artificial Intelligence and Data Authority of India (AIDAI) as an inclusive hub for AI innovation. In addition to NITI Aayog's Operationalizing principles for Responsible AI (2021), MeitY is actively involved in AI policymaking and has proposed the creation of the said AIDAI. MeitY issued an advisory in March 2024, aimed at regulating unreliable AI models and LLMs, directing that platforms using these technologies must ensure compliance with labelling of AI-generated content, AI model to go under test and must not facilitate bias. Additionally, the Telecom Regulatory Authority of India (TRAI) recommended a regulatory framework that would work as an independent statutory authority and a Multi Stakeholder Body (MSB) that would serve as an advisory entity[17]. MeitY is of the opinion that amending the Information technology Act, 2000 would be less time-consuming than adopting new legislation such as the

[15] The Hindu, "G20 New Delhi Leaders' Declaration"; "No Plan to Regulate AI, IT Ministry tells Parliament", April 5, 2023. https://www.thehindu.com/news/national/no-plan-to-regulate-ai-it-ministry-tells-parliament/article66702044.ece

[16] Bharati, Dr. Rahul, *Navigating the Legal Landscape of Artificial Intelligence: Emerging Challenges and Regulatory Framework in India* (July 14, 2024). Available at SSRN: https://ssrn.com/abstract=4898536

[17] PTI, Outlook Business, TRAI *Moots Setting Up Of Statutory Authority Immediately For Development Of Responsible AI, Use Case Regulation,* July 20, 2023. https://www.outlookbusiness.com/news/trai-moots-setting-up-of-statutory-authority-immediately-for-development-of-responsible-ai-use-cases-regulation-news-304602#:~:text=It%20said%20an%20independent%20statutory,of%20use%20cases%20in%20India.

proposed Digital India Act.[18]

From 2023 to 2025, recent judgments on unauthorized AI-generated endorsements, deepfakes and voice cloning illustrate how courts are now slowly interpreting and adapting established doctrines – such as personality rights, negligence, fair use – to address infringement of rights by generative AIs.

The Delhi High Court granted an injunction[19] after AI-driven deepfake videos falsely portrayed Dr. Trehan giving medical advice with his likeness and hospital logo. The court issued a John Doe dynamic injunction, directing intermediaries to remove content and disclose user details within specified timelines. The same court delivered a landmark judgment for the people in the entertainment industry to seek protection against malevolent AI use. Indian actor Anil Kapoor filed a suit seeking protection of his own name, voice likeness, images, persona and other attributes of his personality, from the malevolent AI generated contents. In this case, the defendants used generative AIs to create deepfakes of the actor and sold merchandise or motivational courses by creating false endorsements. Some even used his name, dialogues and voice as ringtones as was reported. Hon'ble Justice Pratibha M. Singh ruled that *"the court cannot turn a blind eye to such misuse of a personality's name and other elements of his persona."* And it is also notable to mention that a first-of-its-kind injunction[20] was issued by the Bombay High Court prohibiting unauthorised AI voice cloning and AI-generated video endorsements of singer Arijit Singh. The court expanded protection to all physical/digital media, including deepfakes, GIFs, avatars and even the metaverse.

CONCLUSION

In an increasingly common scenarios, AI is doing exactly what it was designed to do: choosing its next course of action based on predictive models, routines and feedback. So, if an AI-based trading bot sells off shares minutes before a market crash based on probabilistic models and reasoning, will it be a rational autonomous decision or a dangerous overreach of system? The law, as it stands, does not ascribe intent to machines. But the fact that the system 'chose' to

[18] Aashish Aryan, *"Govt May Amend IT Act to Add New Rules For AI, Genai Models*," Economic Times, January 4, 2024. https://economictimes.indiatimes.com/tech/technology/govt-may-amend-it-act-to-add-new-rules-for-ai-genai-models/articleshow/106524019.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst
[19] Global Health Ltd & Anr v. John Doe & Ors [CS (COMM) 6/2025]
[20] Arijit Singh v. Codible Ventures LLP, (2024) SCC Online Bom 2445

perform that action, raises questions about whether we need new frameworks – one that recognise such intent, by accessing the predictable outputs of AI even when no human intended them.

How do we define the intent when an algorithm makes the decision or worse, how do we differentiate AI-generated content from the real ones?

If an AI credit scoring algorithm starts penalizing applicants based on irrelevant or erroneous data, the blame typically falls on its developers or developing entity. If an algorithm is hacked, or influenced by adversarial data inputs, the system is no longer acting autonomously but under compromised conditions. However, this becomes murky in case AI evolve over time without explicit instructions for each decision they make. Then, can an emergent pattern of bias be considered a technical fault, or does it reflect an unintended, but not wholly unforeseeable, consequence of its design? There is no easy answer to these questions.

India's legal system is still adapting to this new era of technological advancement and thus, it is high time that it not only regulates but also be able to foresee, understand and challenge the complexity and harm before it happens.