CLRS International Conference on Law in the Digital Decade

Law in the Digital Decade

Evidence, Intellectual Property and Markets

Edited by

Gyan Prakash Kesharwani · Prasanna Kumar Shukla

VidhiPragaz® Inking Your Brain

Centre For Legal Research And Studies

Law in the Digital Decade

Evidence, Intellectual Property and Markets

Law in the Digital Decade

Evidence, Intellectual Property and Markets

EDITED BY

Gyan Prakash Kesharwani

Chairperson and Publishing Editor at VidhiAagaz, India

Prasanna Kumar Shukla

Doctoral Researcher at National Law Institute University, Bhopal, Madhya Pradesh, India

IN ASSOCIATION WITH

Centre For Legal Research And Studies

Centre for Legal Research and Studies

VidhiAagaz® Inking Your Brain

MAIHAR · INDIA

VIDHIAAGAZ
Maihar, Madhya Pradesh 485771, India
www.vidhiaagaz.com

Published in association with the Centre for Legal Research and Studies. This book collects papers from the CLRS International Conference on Law in the Digital Decade, held online on 13 September 2026.

© VidhiAagaz 2026
Copyright in each chapter is held as set out in the agreement executed for it.
First published 2026

This book is published open access under the Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0), https://creativecommons.org/licenses/by-nc/4.0/. It may be copied, shared and adapted for non-commercial purposes, provided the authors and the source are credited. Commercial reproduction requires the prior written permission of the publisher.

The views expressed in each chapter are those of its authors. They do not necessarily represent the views of the editors, the Centre for Legal Research and Studies, or the publisher, none of whom accepts responsibility for them.

ISBN 978-81-181981-0-9
DOI https://doi.org/10.63108/VAB.LDD.2

Typeset in Garamond by VidhiAagaz, Maihar
Printed and bound in India

Contents

Preface vii

Notes on Contributors ix

PART I DIGITAL EVIDENCE, CYBERCRIME AND THE CRIMINAL PROCESS

1 Digital Forensic Readiness in India: Rethinking Criminal Investigations in the Digital Age 3
Tamanna Khatri and Dr. Grishma Bhavsar

2 The Certificate Trap: How India's Rule for Electronic Evidence Undermines Cybercrime Prosecutions 15
Vaishali Verma

3 When the Certificate Lies: Rethinking Authenticity Under the Bharatiya Sakshya Adhiniyam 23
Rachel Delfin

4 When Seeing Is No Longer Believing: Deepfakes, Synthetic Evidence and the Criminal Process in India 31
Shirvi Goyal

5 Reading the Guilty Mind: Neurotechnology, Brain Data and the Evidentiary Determination of Mens Rea 41
Debasrita Choudhury

6 Artificial Intelligence Enhanced Forensics and Child Rights Protection: Policy, Law and Ethics 51
Yamini Patel and Dr. Grishma Bhavsar

7 Artificial Abuse: A Socio-Legal Analysis on the Effect of Usage of Generative Artificial Intelligence Against Vulnerable Populations 59
Sharon Thomas and Soorya Narayanan

8 Cybercrime in the Metaverse: Identity, Jurisdiction and Digital Evidence 69
Rupali Mehta and Ashwini Yadav

PART II INTELLECTUAL PROPERTY AND GENERATIVE AI

9 Intellectual Property in the Age of Generative Technology: Rethinking Authorship, Training Data, and Infringement Liability 85
Dr. Shyamal M. Dave and Aditi Patel

10 A Comparative Legal Analysis of the Emerging Right to Know Concerning the Use of Copyrighted Works in AI Training Data 93
C. Sophia Jeyakar

11 Copyright Infringement and Corporate Liability: Navigating the Modern Digital Landscape 109
Dr. Pallavi Pal

12 Protecting Trademark in the Metaverse: The Emerging Threat of AI-Generated Counterfeits 123
A. Vedha Valli

PART III MARKETS, CORPORATIONS AND THE DIGITAL ECONOMY

13 Consumer Protection Against Dark Patterns in E-Commerce: Regulating Manipulative Online Choice 141
Prachi Sharma

14 Dark Patterns in Indian E-Commerce: Emerging Challenges for Consumer Protection and Algorithmic Accountability 149
Parth Sharma

15 The Invisible Employer: Algorithmic Management and the Reinvention of Control in the Gig Economy 157
Chandralekha

16 From Corporate Control to Algorithmic Control: Rethinking Directors' Duties and Liability in AI-Driven Corporations 165
K G Sai Samyukhtha

17 Algorithmic Decision Making and Corporate Insolvency: The Future of Creditor Control in the Digital Decade 177
Ashutosh Mani Pathak

18 Cyber Risk Insurance for Fintech and Digital Payments: Regulatory Challenges and Emerging Coverage Gaps in India 191
Sowmya. H.A

19 The Janus Paradox: Evaluating Tax Friction, Regulatory Blind Spots and Capital Flight Under the Indian Virtual Digital Assets Regime 207
Tarun Kumar Srivastava and Keshav

Preface

Evidence, property and markets occupied the law long before computers did. In the digital decade each has changed shape. A convincing video can be fabricated in minutes, and a certificate can vouch for a record without telling a court whether it is genuine. A machine can produce text, images and music that copyright law was not written for. Platforms can steer what consumers buy and how workers are paid without a human manager in sight.

This book is one of two volumes drawn from the CLRS International Conference on Law in the Digital Decade, which the Centre for Legal Research and Studies at VidhiAagaz held online on 13 September 2026. Its nineteen chapters deal with evidence, intellectual property and markets; the companion volume, Law in the Digital Decade: Rights, Regulation and Accountability, takes up the automated state, the regulation of artificial intelligence, courts, and privacy and speech online.

Part I deals with digital evidence, cybercrime and the criminal process: forensic readiness in criminal investigation, the certificate rule for electronic records under the Bharatiya Sakshya Adhiniyam, 2023, deepfakes as synthetic evidence, what brain data can tell a court about mens rea, AI-enhanced forensics in cases of child sexual abuse, the abuse of generative AI against vulnerable groups, and crime in the metaverse.

Part II turns to intellectual property and generative AI: authorship and infringement when machines produce the work, a right to know whether copyrighted works were used to train a model, the liability of companies for infringement, and the protection of trademarks against AI-generated counterfeits in the metaverse.

Part III considers markets, corporations and the digital economy: dark patterns in e-commerce, algorithmic management of gig workers, directors' duties in corporations that rely on AI, algorithmic decision-making in corporate insolvency, cyber risk insurance for fintech and digital payments, and the taxation and regulation of virtual digital assets.

Every chapter has been read line by line. Where an error of fact, citation or expression could be checked against its source, we have corrected it, and we have tried to leave every argument as its author made it. The views expressed are the authors' own.

We thank the authors, the chairs and co-chairs who led the conference sessions, and the editorial team at VidhiAagaz who prepared the manuscripts for the press. We hope the book will be useful to students, researchers, practitioners and policymakers working where law meets technology.

Gyan Prakash Kesharwani
Chairperson and Publishing Editor at VidhiAagaz, India

Prasanna Kumar Shukla
Doctoral Researcher at National Law Institute University, Bhopal, Madhya Pradesh, India

EDITORS

Notes on Contributors

1. TAMANNA KHATRI, PhD Scholar at Faculty of Law, GLS University, Ahmedabad, Gujarat, India.
   DR. GRISHMA BHAVSAR, Assistant Professor at Faculty of Law, GLS University, Ahmedabad, Gujarat, India.
2. VAISHALI VERMA, Research Scholar at Dr. Ram Manohar Lohiya National Law University, Lucknow, Uttar Pradesh, India.
3. RACHEL DELFIN, Law Student at Christ (Deemed to be University), Bengaluru, Karnataka, India.
4. SHIRVI GOYAL, Student at Law College Dehradun, Uttaranchal University, Dehradun, Uttarakhand, India.
5. DEBASRITA CHOUDHURY, Academician.
6. YAMINI PATEL, PhD Research Scholar at GLS University, Ahmedabad, Gujarat, India.
   DR. GRISHMA BHAVSAR, Assistant Professor at GLS University, Ahmedabad, Gujarat, India.
7. SHARON THOMAS, Student at VIT School of Law, Vellore Institute of Technology, Chennai, Tamil Nadu, India.
   SOORYA NARAYANAN, Student at VIT School of Law, Vellore Institute of Technology, Chennai, Tamil Nadu, India.
8. RUPALI MEHTA, Research Scholar at Department of Law, Panjab University, Chandigarh, India.
   ASHWINI YADAV, Research Scholar at NIILM University, Kaithal, Haryana, India.
9. DR. SHYAMAL M. DAVE, Assistant Professor at Department of Law, Parul Institute of Law, Parul University, Vadodara, Gujarat, India.
   ADITI PATEL, Assistant Professor at Department of Law, Parul Institute of Law, Parul University, Vadodara, Gujarat, India.
10. C. SOPHIA JEYAKAR, Assistant Professor & Research Scholar at Vels Institute of Science, Technology and Advanced Studies, Chennai, Tamil Nadu, India.
11. DR. PALLAVI PAL, Independent Practitioner and Academician.
12. A. VEDHA VALLI, Assistant Professor at Department of Intellectual Property Law, The Tamil Nadu Dr. Ambedkar Law University, Chennai, Tamil Nadu, India.
13. PRACHI SHARMA, Assistant Professor at Vivekananda Institute of Professional Studies, Technical Campus, New Delhi, India.
14. PARTH SHARMA, Research Scholar at National University of Study and Research in Law, Ranchi, Jharkhand, India.
15. CHANDRALEKHA.
16. K G SAI SAMYUKHTHA, VIT School of Law, Vellore Institute of Technology, Chennai, Tamil Nadu, India.
17. ASHUTOSH MANI PATHAK, PhD Research Scholar at National Law University Odisha, Cuttack, Odisha, India.

18. SOWMYA. H.A, PhD Scholar at KLE Law College (A Constituent College of KLE Technological University), Bengaluru, Karnataka, India.

19. TARUN KUMAR SRIVASTAVA, 4th Year B.A. LL.B. (Hons.) Student at Central University of South Bihar, Gaya, Bihar, India.
KESHAV, 4th Year B.A. LL.B. (Hons.) Student at Central University of South Bihar, Gaya, Bihar, India.

PART I

Digital Evidence, Cybercrime and the Criminal Process

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.1

CHAPTER 1

Digital Forensic Readiness in India: Rethinking Criminal Investigations in the Digital Age

TAMANNA KHATRI* AND DR. GRISHMA BHAVSAR**

With the increasing reliance on smartphones, cloud computing technologies, social media sites, mobile money transfer systems, security networks and internet of things (IoT), the modus operandi of criminals along with that of conducting crime investigations, has been transformed to a great extent. Digital evidence has become crucial not only for cyber-crime investigations but also for traditional crimes. Nevertheless, it all depends upon whether the investigation agencies are proficient in handling the digital evidence before it becomes inaccessible through deletion or any other means. In this research, an effort has been made to assess the level of preparedness in digital forensics within the criminal justice process in India. This study aims at determining if the investigating bodies and forensics establishments have the necessary infrastructure, human resources, procedural framework and coordination mechanisms to deal with digital evidence in a timely and legal manner. The research methodology employed in this research is doctrinal and analytical. The current paper draws a distinction between digital forensic readiness and traditional digital forensics. In particular, whereas forensic investigation usually starts after the occurrence of a crime, forensic readiness implies advance preparation through retention of evidence, secure logging, first response training, incident response strategies and proper forensic labs. The main problems associated with digital forensic readiness in India include unequal distribution of infrastructure, lack of specialized experts, poor first response training, delay in forensic investigation and challenges in getting information from foreign and private service providers. This paper emphasizes that the mere legal recognition of electronic evidence is insufficient without any other preparations. It suggests creating a strategy for digital forensic readiness that relies on standardized procedures, professional training, accreditation of laboratories, secure evidence management system and collaboration of the criminal justice system.

Keywords: Digital Forensic Readiness, Digital Evidence, Criminal Investigation, Forensic Infrastructure, Institutional Preparedness, Criminal Justice System.

1 INTRODUCTION

The increased reliance on digital technologies has transformed the way crimes are committed, detected and investigated. Many of today's human activities leave behind a digital footprint. Mobile phone communications, financial transactions done online, social media communications, location tracking services, online cloud computing, closed-circuit television systems and other connected devices can all produce information which might become important for the purposes of criminal investigation. As a consequence of this development, digital evidence has ceased being confined to the category of cybercrimes. Digital evidence may be useful in cases of homicide, kidnapping, organized crime, financial crimes, drug trafficking, corruption, sexual offences and many other traditional crimes.[1] These developments have put

* PhD Scholar at Faculty of Law, GLS University, Ahmedabad, Gujarat, India.
** Assistant Professor at Faculty of Law, GLS University, Ahmedabad, Gujarat, India.

[1] UNITED NATIONS OFFICE ON DRUGS AND CRIME, DIGITAL EVIDENCE, CYBERCRIME MODULE 4; INTERPOL, METAVERSE: A LAW ENFORCEMENT PERSPECTIVE—USE CASES, CRIME, FORENSICS, INVESTIGATION, AND GOVERNANCE.

additional obligations on the investigative agencies. The mere presence of digital evidence does not mean it will be useful in the course of an investigation. The usefulness of digital evidence will depend on its timely identification, its preservation in its original state, proper acquisition techniques and finally analysis by the qualified specialists. Electronic evidence can be overwritten, deleted, modified remotely, encrypted and stored on devices outside the control of the investigating agency. In other words, the success of any digital investigation depends on the pre-investigatory work done in advance.[2]

This is how the idea of digital forensic readiness emerges. In general terms, forensic readiness can be understood as an institution's/forensic investigation system's capacity to foresee the possibility of using the digital evidence and develop the necessary systems for collecting and using it. This idea draws attention from the reactive model of forensic investigation towards a more systematic one.[3] Similarly, international forensic guidelines have underlined the importance of incorporating forensics into incident response procedures instead of carrying out the forensic examination independently.[4] India has introduced legislative measures regarding the use of electronic evidence. According to the Bharatiya Sakshya Adhiniyam, 2023, electronic or digital evidence is allowed to be used and such evidence cannot be refused admission just on account of its being stored in the electronic form. The Information Technology Act, 2000 also allows the central government to designate government departments or bodies as Examiners of Electronic Evidence.[5]

In addition to this, the Bharatiya Nagarik Suraksha Sanhita, 2023 has added further technological components to criminal procedure like the recording of searches and seizures by audio-visual means of electronics and processes of seeking help in investigation with regard to foreign territories.[6] All these factors clearly prove that digitization has now become an integral part of the criminal justice process. But then, mere legalization of digitalization and procedural reform do not guarantee forensic readiness. It depends also on the preparedness of the officers of police, cyber-crime unit, forensic labs, prosecutor and other agencies. Hence, the issue is not just limited to the permissibility of the use of digital evidence in the Indian law, but the preparedness of the concerned authorities to obtain and process such evidence. The present paper attempts to analyze this issue in a doctrinal way. It defines the concept of forensic readiness with respect to digitalization, differentiates it from traditional digital forensics, looks at the present Indian institutional and legal framework, discusses the major problems involved in this issue and suggests measures to develop forensic readiness.

2 UNDERSTANDING DIGITAL FORENSIC READINESS

Digital forensics usually implies a series of procedures by which information kept or sent via electronic means is found, collected, acquired, retained, analysed and interpreted for investigative purposes. Standards like ISO/IEC 27037 offer recommendations related to identification, collection, acquisition and

[2] INT'L ORG. FOR STANDARDIZATION, ISO/IEC 27037:2012, INFORMATION TECHNOLOGY—SECURITY TECHNIQUES—GUIDELINES FOR IDENTIFICATION, COLLECTION, ACQUISITION AND PRESERVATION OF DIGITAL EVIDENCE (2012).

[3] Robert Rowlingson, *A Ten Step Process for Forensic Readiness*, 2 INT'L J. DIGIT. EVID. 1 (2004).

[4] KAREN KENT, SUZANNE CHEVALIER, TIM GRANCE & HUNG DANG, GUIDE TO INTEGRATING FORENSIC TECHNIQUES INTO INCIDENT RESPONSE, NIST Special Publication 800-86 (2006).

[5] Bharatiya Sakshya Adhiniyam, 2023, § 61; Information Technology Act, 2000, § 79A.

[6] Bharatiya Nagarik Suraksha Sanhita, 2023, §§ 105, 112–13.

retention of digital evidence.[7] However, digital forensic readiness is a slightly different issue. It has to do with making sure that organizations are prepared for a situation when digital information could become evidence at some point in the future. More precisely, while traditional digital forensics seeks to find out how evidence could be analysed after a problem arose, forensic readiness focuses on how well the organization was prepared to retain evidence in the first place.[8] The significance of this is that digital evidence has a limited life span. Logs could be auto-erased from the systems, cloud storage information would have been saved for limited periods, CCTV footage could be erased once space constraints are achieved and information stored on personal devices could have been tampered with during regular usage. If pertinent information is not captured at the right time, it will later be difficult for a sophisticated forensic laboratory to recreate it.[9] Preparedness for forensics is more than just the availability of forensic tools. It entails having the infrastructure, skilled staff, roles and responsibilities, documentation process and methods of retaining evidence.[10] At the level of the investigation, preparedness starts with the first officer that comes across any digital device or digital environment which could possibly provide relevant information. That officer should have the ability to identify whether the information available in the mobile phone, computer, security cameras, etc., is relevant and know how to preserve it until more specialized help arrives.[11] At the organizational level, preparedness refers to the ability of the forensic laboratory to collect, preserve, analyse and provide information about the digital evidence within an appropriate time frame. This also includes collaboration between investigators, forensics, prosecutors and technology service providers. Thus, the idea denotes the transition from a forensic to a preparedness-based investigation approach.[12]

3 IMPORTANCE OF DIGITAL FORENSIC READINESS IN MODERN CRIMINAL INVESTIGATION

Forensic readiness becomes necessary due to the evolving nature of evidence. Physical evidence is always in some tangible form that can be segregated from its surroundings. On the other hand, digital evidence is usually present in multiple devices, networks, platforms and jurisdictions. Hence, the process of its identification and acquisition requires investigators to know not just the physical device, but also the larger digital environment where the evidence is stored.[13] One investigation might involve information from the mobile phone of the accused, as well as information from CCTV cameras, social media websites, cloud-based storage, financial institutions, telecommunications providers and third-party applications. The information from all of these sources is not controlled by the same agency and might not even be located in the jurisdiction conducting the investigation. Therefore, forensic readiness enables the

[7] INT'L ORG. FOR STANDARDIZATION, ISO/IEC 27037:2012, INFORMATION TECHNOLOGY—SECURITY TECHNIQUES—GUIDELINES FOR IDENTIFICATION, COLLECTION, ACQUISITION AND PRESERVATION OF DIGITAL EVIDENCE (2012).
[8] Robert Rowlingson, *A Ten Step Process for Forensic Readiness*, 2 INT'L J. DIGIT. EVID. 1 (2004).
[9] BARBARA GUTTMAN, DOUGLAS R. WHITE & TRACY WALRAVEN, DIGITAL EVIDENCE PRESERVATION: CONSIDERATIONS FOR EVIDENCE HANDLERS, NISTIR 8387 (2022).
[10] Robert Rowlingson, *A Ten Step Process for Forensic Readiness*, 2 INT'L J. DIGIT. EVID. 1 (2004).
[11] NAT'L INST. OF JUST., U.S. DEP'T OF JUST., ELECTRONIC CRIME SCENE INVESTIGATION: A GUIDE FOR FIRST RESPONDERS (2d ed. 2008).
[12] KAREN KENT ET AL., GUIDE TO INTEGRATING FORENSIC TECHNIQUES INTO INCIDENT RESPONSE, NIST Special Publication 800-86 (2006).
[13] INT'L ORG. FOR STANDARDIZATION, ISO/IEC 27037:2012, GUIDELINES FOR IDENTIFICATION, COLLECTION, ACQUISITION AND PRESERVATION OF DIGITAL EVIDENCE.

investigator to identify such sources right at the beginning rather than confining the investigation to the device seized at the crime scene.[14]

The importance of preserving such information is more significant in cases where evidence is vulnerable and time sensitive. Logs could have been overwritten; cloud data could have been saved for limited times; CCTV footage could have been automatically deleted and there could be cases whereby the data changes through use of the device. Thus, delay would lead to loss of information, change in metadata or expiry of retention periods. Therefore, digital investigations are greatly dependent on the speed at which relevant sources are identified and preservation measures taken.[15] Forensic readiness is directly proportional to the integrity and reliability of evidence collected in a case. Forensic readiness makes it easy for investigators to put in place standard procedures for seizing, documenting, forensic imaging, storage and transfer of digital devices. These standard procedures reduce unnecessary handling of electronic evidence and make it easy to establish an undisturbed chain of custody. The chain of custody is very crucial especially when the authenticity of the electronic evidence collected is challenged.[16]

The second key factor is investigative efficiency. When the investigators know what kind of evidence could possibly exist in the crime scene and how it needs to be handled, the forensic laboratories get much better-quality evidence for their analysis. Proper standardisation will also help to eliminate unnecessary delays that can occur due to inadequate documentation or failure of seizure or loss of evidence prior to its specialist analysis. Thus, forensic readiness allows using available technical capabilities more efficiently.[17] It is important from the standpoint of forensic readiness that the digital forensic examination does not start when the evidence arrives at the specialized laboratory. It should start right from the first investigator who comes into contact with the digital crime scene. Forensic readiness can have a major impact on the usefulness of evidence.[18] It follows that forensic readiness should not be considered just from a cybersecurity standpoint. In the criminal justice setting, forensic readiness is very much linked with investigative efficiency, evidentiary soundness and organizational competence. A legal system which accepts digital evidence but has not made the preparations to seize and secure the evidence could actually fail to gather crucial evidence for forensic analysis.

4 LEGAL FOUNDATION FOR DIGITAL EVIDENCE IN INDIA

The legal framework in India has increasingly acknowledged the role of electronic and digital records as a valid mode of evidence. The Bharatiya Sakshya Adhiniyam, 2023, which commenced from 1st July 2024, addresses the issue of the law of evidence in general in India. Sections 61 to 63 are particularly relevant for discussing electronic or digital records and their admissibility as evidence. While Section 61 defines that an electronic or digital record shall not be regarded as inadmissible simply on account of being in electronic form, Section 63 lays down the condition of computer outputs and electronic

[14] KAREN KENT, SUZANNE CHEVALIER, TIM GRANCE & HUNG DANG, GUIDE TO INTEGRATING FORENSIC TECHNIQUES INTO INCIDENT RESPONSE, NIST Special Publication 800-86 (2006).
[15] *Id.*; INT'L ORG. FOR STANDARDIZATION, ISO/IEC 27037:2012.
[16] INT'L ORG. FOR STANDARDIZATION, ISO/IEC 27037:2012, GUIDELINES FOR IDENTIFICATION, COLLECTION, ACQUISITION AND PRESERVATION OF DIGITAL EVIDENCE.
[17] KAREN KENT, SUZANNE CHEVALIER, TIM GRANCE & HUNG DANG, GUIDE TO INTEGRATING FORENSIC TECHNIQUES INTO INCIDENT RESPONSE, NIST Special Publication 800-86 (2006).
[18] *Id.*; ERAN SALFATI & MICHAEL PEASE, DIGITAL FORENSICS AND INCIDENT RESPONSE FRAMEWORK FOR OPERATIONAL TECHNOLOGY, NISTIR 8428 (2022).

records.[19] There is also provision in the legislation regarding the involvement of experts. Section 39 of the Bharatiya Sakshya Adhiniyam is concerned with expert opinion and provides for relying on persons having special knowledge in relevant subjects. Furthermore, Section 79A of the Information Technology Act, 2000, gives power to the Central Government to notify an appropriate government department, body or agency as an Examiner of Electronic Evidence to give expert opinion in courts and other authorities.[20] This becomes particularly significant in cases in which the validation, acquisition or interpretation of electronic evidence requires particular forensic expertise. The application of such a technique is particularly significant in the context of forensic readiness in that it allows one to develop an organizational framework within which specialized laboratories can contribute to judicial proceedings in a formal manner. For instance, the Directorate of Forensic Science, Gandhinagar, Gujarat, was notified as an Examiner of Electronic Evidence under Section 79A in 2018.[21]

The procedural regime has become technology-centric as well. Under Section 105 of the Bharatiya Nagarik Suraksha Sanhita, 2023, the process of conducting search and seizure shall be recorded electronically using audio-visual electronic method, preferably using a mobile phone.[22] Such recording will help achieve more transparency and proper documentation of the investigative process. The legislation has mechanisms dealing with letters of request and cooperation in investigations where the evidence or person is located in foreign territories.[23] All in all, such provisions indicate greater acceptance of technology in the fields of evidence laws and criminal procedure regime. The law has accepted the use of electronic records, specialist expertise and technological aid in investigation. However, such provisions ensure only the creation of legal authorization and safeguard of evidence, not necessarily the ability to handle digital evidence. This dichotomy plays an important role in the idea of forensic readiness. The law can allow for the collection and use of electronic evidence, yet readiness will determine if that evidence is able to be found, collected, analyzed and presented. This means that a sound legal basis needs to be paired with forensic capabilities in order for digital evidence to reach its full potential.

5 Institutional Developments for Digital Forensic Investigation in India

A lot of institutional developments have taken place in India, which aim at improving the capacity of the state to investigate cyber-crimes through digital forensic investigation. The term "institutional development" means that a state tries to develop infrastructure and process of coordination that would enable it to handle digital evidence in criminal investigations. For instance, the Ministry of Home Affairs formed the Indian Cyber Crime Coordination Centre (I4C) which serves as a coordinating body to handle cybercrimes within India. The institutional bodies include the National Cybercrime Threat Analytics Unit, National Cybercrime Reporting Portal, Platform for Joint Cybercrime Investigation Teams, National Cybercrime Forensic Laboratory (NCFL) Ecosystem, National Cybercrime Training Centre, Cybercrime Ecosystem Management Unit, and National Cyber Crime Research and Innovation Centre.[24] The NCFL Ecosystem is particularly important in the context of forensic readiness since it seeks to support forensic

[19] Bharatiya Sakshya Adhiniyam, 2023, §§ 61–63.
[20] Bharatiya Sakshya Adhiniyam, 2023, § 39; Information Technology Act, 2000, § 79A.
[21] Ministry of Electronics and Information Technology, Government of India, Notification under § 79A of the Information Technology Act, 2000, notifying the Directorate of Forensic Science, Gandhinagar, Gujarat as an Examiner of Electronic Evidence (2018).
[22] Bharatiya Nagarik Suraksha Sanhita, 2023, § 105.
[23] *Id.* §§ 112–113.
[24] Ministry of Home Aff., Gov't of India, Indian Cyber Crime Coordination Centre (I4C) Scheme.

investigations using innovative technologies. According to the I4C, adequate facilities with the necessary equipment and sufficient numbers of skilled personnel should be available in order to keep up with technological advancements and support investigations.[25] These specialized facilities will help in carrying out investigations where the investigating officers do not have the required technical capabilities to conduct the examination of evidence collected.

In addition, the Ministry of Home Affairs has assisted in setting up cyber forensic and training laboratories in the States and Union Territories, besides conducting skill enhancement programs for law enforcement personnel, prosecuting officers and judicial officers. By June 2026, there were 33 laboratories set up in States and Union Territories, while over 24,600 law enforcement officers, prosecuting officers and judicial officers were trained in various subjects including cybercrime investigation and forensics.[26] It is especially pertinent because forensic readiness is highly dependent on human resources as well. More efforts have been devoted to developing forensics infrastructure at the State level. The government's initiatives aimed at improving DNA and cyber forensics have helped improve cyber-forensic infrastructure in State Forensic Science Laboratories. As mentioned by the Ministry of Home Affairs, the assistance provided through such initiatives has enabled the upgrade of DNA and cyber-forensic laboratories in State FSLs, with the aim of improving scientific examination in criminal investigation cases.[27]

The described steps show that India has achieved considerable progress in building its institutional digital forensics capacity. Creation of specialized laboratories, training programs and national coordination systems is very important for building forensic readiness. Nevertheless, forensic readiness is not only about national laboratories or cybercrime units. Relevant capacity must be available at the stage of the first investigation to ensure proper preservation and documentation of digital evidence. The problem here then lies in making sure that the specialised institutional capabilities are converted into effective operational readiness at various levels of the criminal justice process. The national level infrastructure will be able to offer expertise and advanced forensic capability; however, the success of such an effort is contingent on collaboration with State authorities and the local authorities at the frontline.

6 MAJOR CHALLENGES TO DIGITAL FORENSIC READINESS IN INDIA

6.1 Inequalities in Institutional Capacities

Among the main challenges that can arise when trying to create forensic readiness is the issue of creating equal capacity in different geographical and institutional environments. Digital evidence can occur in any police station and in any criminal investigation, and not only in the specialized departments for dealing with cybercrime. This means that forensic knowledge, adequate tools, and specialists must be available in all large urban centres and not only in specialized institutions. A high-tech forensic lab can give an advanced forensic analysis, but its efficiency depends on how good the evidence brought to the lab is. If the evidence was poorly handled, documented, or preserved at the beginning of the process, it can considerably affect the whole forensic investigation. Acknowledging the significance of making forensic readiness available to more people, parliamentary committees have pointed out the need to improve forensic

[25] MINISTRY OF HOME AFF., GOV'T OF INDIA, NATIONAL CYBERCRIME FORENSIC LABORATORY ECOSYSTEM, I4C SCHEME.
[26] PRESS INFO. BUREAU, GOV'T OF INDIA, MINISTRY OF HOME AFF., NATIONAL CYBER CRIME DATA (July 21, 2026).
[27] MINISTRY OF HOME AFF., GOV'T OF INDIA, STRENGTHENING OF DNA ANALYSIS AND CYBER FORENSIC CAPACITIES IN STATE FSLS (updated Feb. 20, 2026).

infrastructure and use mobile forensic labs.[28] Forensic readiness should thus not be limited to national or State institutions, but rather reach the local investigating officer. The efficiency of advanced forensic infrastructure ultimately relies on readiness from the point where digital evidence is discovered.

6.2 First Responder Training

A critical role is played by the first responder coming across any digital evidence. Digital devices cannot always be treated like regular physical exhibits, as the interaction with the device might change some of the potential evidence. Interaction with an unlocked mobile phone, for instance, may change system data. Unplugging or turning off the specific devices without knowing all the facts may lead to loss of volatile data. Likewise, not detecting other connected accounts, networks, or cloud storage services may cause investigators to look at only the physical device and miss the remote location evidence. It is crucial, therefore, to have the correct first response procedure in place to make sure the value of digital evidence remains intact. It is unrealistic to train every police officer to become a forensic scientist. On the contrary, the investigating officers need enough knowledge of how to detect digital evidence, secure and document the scene, preserve potential evidence and seek help from specialists where needed. In this regard, the Bureau of Police Research and Development itself has specialized courses on cybercrime investigation, mobile forensics, and crime scene investigation.[29]

6.3 Lack of Specialized Expertise

Forensic analysis is becoming more and more specialized. Current cases could involve mobile operating systems, encrypted apps, cloud computing, cryptocurrencies, Internet-of-things devices, malware, or a lot of multimedia data. Every one of these areas might demand its own tools, methods, and type of technical knowledge. It is difficult to possess expertise in all these areas due to the continuous evolution of technologies, operating systems, and forensic tools. Knowledge obtained during training could, thus, become obsolete due to the emergence of new technologies, platforms and ways of information storage or concealment. Forensic preparedness calls for continuous professional development and regular updating of technical skills. The very approach of the Indian institutions demonstrates awareness of this need. The National Cybercrime Training Centre within I4C aims at standardized training in cybercrimes, investigative training and training in simulated cyber environment, including a proposal of Cyber Range.[30] The Ministry of Home Affairs in India realizes the necessity to train forensic manpower and improve the existing forensic infrastructure.[31]

---

[28] DEP'T-RELATED PARLIAMENTARY STANDING COMM. ON HOME AFF., RAJYA SABHA, REPORT NO. 252, DEMANDS FOR GRANTS (2025-2026), MINISTRY OF HOME AFFAIRS ¶¶ 4.14.6–4.14.7 (Mar. 10, 2025); Ministry of Home Affairs, Government of India, Reply to Lok Sabha Unstarred Question No. 3452, Infrastructure in Forensic Laboratories (Dec. 17, 2024).

[29] BUREAU OF POLICE RSCH. & DEV., GOV'T OF INDIA, TRAINING MATERIAL AND COURSES ON INVESTIGATION OF CYBER CRIME, MOBILE FORENSICS AND CRIME SCENE INVESTIGATION.

[30] MINISTRY OF HOME AFF., GOV'T OF INDIA, INDIAN CYBER CRIME COORDINATION CENTRE—NATIONAL CYBERCRIME TRAINING CENTRE.

[31] Ministry of Home Affairs, Government of India, Reply to Lok Sabha Unstarred Question No. 3452, Infrastructure in Forensic Laboratories (Dec. 17, 2024).

7 DIGITAL EVIDENCE STORED WITH PRIVATE SERVICE PROVIDERS

In modern-day investigations, evidence might not always be physically present in the possession of the person under investigation or even in the jurisdiction where the investigation is being conducted. Such evidence could be available with telecommunications companies, social media providers, cloud storage providers, or any other technology companies. In cases like these, while investigators might have physical possession of a device, the information, which could include account information, communication information, cloud storage backup, or transaction information, could be under the possession of the third-party service provider.[32] It thus becomes a major challenge in terms of forensic readiness as the investigators would need to rely on the third-party service provider to preserve the information and provide the information at a later stage. It becomes imperative for the investigators to be able to figure out who the third-party service provider is, the type of information that is available from that service provider and how preservation or disclosure can be done in a legal manner.[33]

Cooperation is particularly crucial when the data storage or control happens out of India. As a framework of comparative international law, the Budapest Convention on Cybercrime incorporates procedures and international cooperation provisions related to cybercrime and electronic evidence. In addition, the Budapest Convention serves as a model or guideline for international development of a country's approach to the use of electronic evidence and the Second Additional Protocol to it covers enhanced cooperation and disclosure of electronic evidence.[34] The criminal procedure law in India provides for the procedure for the seeking of assistance in investigations from other countries. Sections 112 and 113 of the Bharatiya Nagarik Suraksha Sanhita, 2023 deal with letters of request for investigation in a country or place outside India and letters of request received from foreign countries or places for investigation in India, respectively.[35] However, conventional processes of international assistance are not always as swift as the life cycle of digital evidence. The evidence in question might be fleeting, scattered over different jurisdictions or held by service providers situated in other countries. For this reason, current international talks regarding electronic evidence tend to address rapid evidence preservation, cooperation with service providers and efficient means of access from abroad.[36] Preparedness in the field of forensics implies, among other things, knowledge not only about the way digital evidence needs to be preserved technically, but also about its whereabouts, its controllers and the procedure according to which it needs to be preserved. Identification of service providers in good time, prompt making of preservation requests and knowledge about cooperation procedures at home and abroad can greatly minimize the chances of losing the evidence.

---

[32] COUNCIL OF EUROPE, EFFECTIVE ACCESS TO ELECTRONIC EVIDENCE; COUNCIL OF EUROPE, COVID-19-RELATED CYBERCRIME AND ELECTRONIC EVIDENCE IN ASIA (2022).

[33] Council of Europe, Convention on Cybercrime (Budapest Convention) arts. 16–18, E.T.S. No. 185 (2001); CYBERCRIME CONVENTION COMM. (T-CY), T-CY GUIDANCE NOTE #10: PRODUCTION ORDERS FOR SUBSCRIBER INFORMATION (ARTICLE 18 BUDAPEST CONVENTION), T-CY(2015)16 (adopted Feb. 28, 2017).

[34] Council of Europe, Convention on Cybercrime (Budapest Convention); Second Additional Protocol to the Convention on Cybercrime on Enhanced Co-operation and Disclosure of Electronic Evidence, CETS No. 224.

[35] Bharatiya Nagarik Suraksha Sanhita, 2023, §§ 112–113.

[36] COUNCIL OF EUROPE, INTERNATIONAL COOPERATION ON CYBERCRIME AND ELECTRONIC EVIDENCE; Second Additional Protocol to the Budapest Convention on Cybercrime.

8 INTERNATIONAL STANDARDS AND THEIR SIGNIFICANCE FOR INDIA

Any good forensic readiness approach should take note of international standards on digital evidence collection and analysis. ISO/IEC 27037 is a standard providing recommendations concerning the identification, collection, acquisition and preservation of possible digital evidence. This includes a wide variety of evidence sources including computers, mobile devices, digital cameras, CCTV systems, networks and related technologies.[37] Also, NIST Special Publication 800-86 focuses on incorporating forensics into the incident response process. Instead of treating forensics as an independent and last resort procedure, this guide recommends understanding available data sources and developing approaches to conduct the proper forensics. Another NIST guidance about digital forensics and incident response regarding operational technology underlines the importance of preparation for incidents in advance, including formation of incident response teams and respective procedures.[38] Such an approach is significant for criminal investigations in India as it focuses on planning, standardization and forensics preparation instead of simple laboratory examination. India is not obligated to adopt all the international standards as they were designed for different legal frameworks. However, the ideas of documentation, responsibility distribution, data preservation, personnel and procedures validation can serve as an inspiration for the country.

9 TOWARDS A DIGITAL FORENSIC READINESS FRAMEWORK FOR INDIA

In order to ensure digital forensic readiness, it is necessary to establish such readiness as an integral criminal justice policy rather than an isolated technical approach. This involves integrating investigation, evidence collection, forensic analysis, prosecution and cooperation between institutions in order to ensure efficient management of digital evidence through its entire life cycle. First of all, there must be standard procedures developed for the first responders to follow. The investigating officers need to be provided with guidance regarding the identification of digital evidence, securing of devices, documentation of electronic scene, and acquisition of specialists' help. Such procedures should also define responsibilities for the first responder to prevent destruction of evidence before forensic analysis starts.[39] Secondly, it is vital to have continuous capacity building. Digital evidence awareness should be included into the usual training, whereas the specialists should undergo additional training which takes into account rapid change of technology. The National Cybercrime Training Centre within the I4C already shows the importance of training.[40]

Thirdly, there needs to be more focus on laboratory capacity and quality assurance. Under Section 79A of the Information Technology Act, 2000, provision has been made for formally notifying Government Departments, bodies or agencies as Examiners of Electronic Evidence so that they may give expert opinions.[41] It is therefore imperative that the necessary laboratory infrastructure is backed up with

---

[37] INT'L ORG. FOR STANDARDIZATION, ISO/IEC 27037:2012, INFORMATION TECHNOLOGY—SECURITY TECHNIQUES—GUIDELINES FOR IDENTIFICATION, COLLECTION, ACQUISITION AND PRESERVATION OF DIGITAL EVIDENCE (2012).

[38] KAREN KENT, SUZANNE CHEVALIER, TIM GRANCE & HUNG DANG, GUIDE TO INTEGRATING FORENSIC TECHNIQUES INTO INCIDENT RESPONSE, NIST Special Publication 800-86 (2006); ERAN SALFATI & MICHAEL PEASE, DIGITAL FORENSICS AND INCIDENT RESPONSE (DFIR) FRAMEWORK FOR OPERATIONAL TECHNOLOGY (OT), NISTIR 8428 (2022).

[39] INT'L ORG. FOR STANDARDIZATION, ISO/IEC 27037:2012, INFORMATION TECHNOLOGY—SECURITY TECHNIQUES—GUIDELINES FOR IDENTIFICATION, COLLECTION, ACQUISITION AND PRESERVATION OF DIGITAL EVIDENCE (2012).

[40] MINISTRY OF HOME AFF., GOV'T OF INDIA, INDIAN CYBER CRIME COORDINATION CENTRE (I4C) SCHEME, NATIONAL CYBERCRIME TRAINING CENTRE.

[41] Information Technology Act, 2000, § 79A.

qualified personnel, valid forensic processes and quality control measures. Fourthly, there is a need to have better systems of managing digital evidence. Digital evidence should be properly managed right from the process of seizure to examination and finally production before the court. The use of access controls, audit trails, chain-of-custody and standardization can increase accountability and reduce disputes about the integrity or management of electronic evidence. Even international digital-evidence standards emphasize the systematic identification, collection, acquisition and preservation of digital evidence.[42] Fifthly, forensic preparedness should facilitate coordination among agencies. Investigating agencies, forensic labs, prosecution and cybercrime units cannot operate in silos. The I4C framework has already incorporated this in its architecture with various mechanisms for joint cybercrime investigation, forensic support, training, research and coordination among law enforcement agencies and other stakeholders.[43]

Sixth, there should be more effective mechanisms for working with private technology companies and handling cross-border digital evidence. Police should be trained not only in handling digital evidence but also in following the processes of collecting digital evidence through legal means. Sections 112 and 113 of the Bharatiya Nagarik Suraksha Sanhita, 2023 give mechanisms concerning letters of request for investigations pertaining to a foreign jurisdiction.[44] Other international conventions also recognize the need for urgent preservation and collaboration when it comes to cross-border electronic evidence. Lastly, readiness should be seen as a continuous process and not just an infrastructure development project. There will always be new types of devices, new platforms, storage devices, and problems related to investigation. Procedures, forensic methods, and professional training therefore need constant update to remain capable.[45]

10 CONCLUSION

Digital evidence has become an ordinary and growing aspect of modern criminal investigations. The key issue before the criminal justice process then is not just whether the electronic evidence is legally recognized, but whether it is properly prepared to discover, secure, analyze and employ it. Under these circumstances, digital forensic readiness becomes the move from a reactive method to an environment where investigative agencies anticipate their need for digital evidence and prepare for it. India is well on its way in this regard. Not only is the Bharatiya Sakshya Adhiniyam, 2023 explicit about electronic and digital documents, the Information Technology Act makes provision for the appointment of Examiners of Electronic Evidence while the Bharatiya Nagarik Suraksha Sanhita, 2023 recognizes technology in criminal proceedings. There have been national initiatives such as I4C, cyber forensics labs and training centers among others. However, the last remaining task is to implement these laws and institutions as consistent readiness throughout the entire process of criminal investigations. Digital forensic readiness cannot be limited to the availability of sophisticated laboratories at institutional levels. Instead, digital forensic readiness must start from the very beginning with the officer in charge of investigating a case to

[42] INT'L ORG. FOR STANDARDIZATION, ISO/IEC 27037:2012, INFORMATION TECHNOLOGY—SECURITY TECHNIQUES—GUIDELINES FOR IDENTIFICATION, COLLECTION, ACQUISITION AND PRESERVATION OF DIGITAL EVIDENCE (2012).

[43] MINISTRY OF HOME AFF., GOV'T OF INDIA, INDIAN CYBER CRIME COORDINATION CENTRE (I4C) SCHEME.

[44] Bharatiya Nagarik Suraksha Sanhita, 2023, §§ 112–113; COUNCIL OF EUROPE, INTERNATIONAL COOPERATION AGAINST CYBERCRIME.

[45] KAREN KENT, SUZANNE CHEVALIER, TIM GRANCE & HUNG DANG, GUIDE TO INTEGRATING FORENSIC TECHNIQUES INTO INCIDENT RESPONSE, NIST Special Publication 800-86 (2006).

include the processes of digital evidence recognition, collection, examination, interpretation, prosecution, and finally presentation in court. At any single stage of the process, weaknesses can render important digital evidence practically useless. In this light, there needs to be a national approach towards forensic readiness which includes standard operating procedures, training programs, laboratory facilities, good digital evidence management systems, inter-agency collaboration and effective measures for acquiring data owned by foreign and privately held service providers. International standards on information security management like ISO/IEC 27037 and forensic guidelines from NIST can help establish useful principles for evidence preservation, documentation, preparation, and incident response in the Indian context.

Forensic readiness, on the other hand, is a continuous process and should not be considered as a one-off technological or infrastructural investment. Digital technologies, storage mediums, means and modes of communications and modes of committing crimes are advancing at a very fast pace. It follows that forensic practices, techniques and skills have to be continuously assessed and updated to ensure that criminal justice institutions are able to respond to these changing modes of digital evidence. The general point is that technology alone cannot improve the efficiency of criminal investigation. Digital evidence will become effective only if the institutions charged with its handling are ready to act in the appropriate manner and at the appropriate time. The development of an organized process of digital forensic readiness will thus be an important component of the ongoing efforts of India to create an efficient and effective criminal justice system.

* * *

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.2

CHAPTER 2

The Certificate Trap: How India's Rule for Electronic Evidence Undermines Cybercrime Prosecutions

VAISHALI VERMA*

Cybercrime prosecutions depend on electronic evidence, yet before a court can weigh whether such evidence is convincing, it must decide whether the evidence can be looked at at all. This paper compares how India, the United Kingdom, and the United States answer that threshold question of admissibility, and argues that India has chosen the least workable of the three answers. India treats a signed certificate under Section 63 of the Bharatiya Sakshya Adhiniyam, 2023 as an absolute precondition to admitting any electronic record, so that genuine evidence can be excluded over paperwork errors unrelated to its reliability. The United Kingdom abandoned a similar certificate requirement in 1999 in favour of judicial discretion, while the United States treats certification as only one of several optional routes to authentication. Drawing on this comparison, the paper argues that India should convert its certificate from a gatekeeper into a factor courts weigh, and proposes concrete statutory and administrative reforms.

Keywords: Electronic Evidence, Cybercrime, Admissibility, Comparative Criminal Law, Digital Evidence Certification

1 INTRODUCTION

Cybercrime investigations run on digital proof. A hacking case rests on server logs. A phishing case rests on emails. An online fraud case rests on bank transaction records and chat histories. Take away the electronic record, and there is often no case left at all. Cybercrime offences in India, from unauthorised access to identity theft, are themselves defined by reference to computers and communication devices under the Information Technology Act, 2000, so the proof of these offences is almost always electronic by definition.[1] Yet before any court can decide whether such evidence is convincing, it must first decide whether the evidence can be looked at in the first place. This is the law of admissibility, and it sits quietly behind every cybercrime trial, deciding cases long before anyone gets to argue about guilt.

This paper compares how three countries, India, the United Kingdom, and the United States, answer one narrow but consequential question: what must a party show before a court will even look at an electronic record. It argues that India has chosen the least workable of the three answers. India insists on a specific signed certificate before any electronic record can be admitted, and treats the certificate as a condition precedent, meaning that without it, the evidence simply cannot be considered, no matter how obviously genuine it is.[2] The United Kingdom abandoned an almost identical certificate requirement decades ago and now lets judges assess reliability case by case.[3] The United States never had a mandatory

* Research Scholar at Dr. Ram Manohar Lohiya National Law University, Lucknow, Uttar Pradesh, India.

[1] Information Technology Act, 2000.

[2] Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1.

[3] Youth Justice and Criminal Evidence Act 1999, c. 23, § 60 (UK).

certificate rule; a certificate is available as a shortcut, but it is only one of several ways to get electronic evidence admitted.[4] This comparison forms part of a broader study of how different legal systems weigh electronic evidence in cybercrime prosecutions, and the certificate question is where the three systems diverge most sharply.

The argument proceeds in six parts. Part 2 explains briefly why electronic evidence needs any special rule at all. Part 3 traces India's certificate requirement from its statutory origin through the Supreme Court's unsettled case law to its current form under the Bharatiya Sakshya Adhiniyam, 2023. Part 4 examines the United Kingdom's move away from certification toward judicial discretion. Part 5 examines the United States' multi-track system of authentication. Part 6 argues that India's all-or-nothing rule, designed to guarantee reliable evidence, has instead become a technical escape route in cybercrime prosecutions, so that genuine evidence gets thrown out over paperwork while prosecutions collapse on procedure rather than merit. Part 7 proposes that India convert the certificate from an absolute gatekeeper into a factor that courts weigh, drawing on the UK and US models, while preserving the certificate's genuine value as a safe harbour.

2 WHY ELECTRONIC EVIDENCE NEEDED A SPECIAL RULE IN THE FIRST PLACE

Paper records are hard to alter without leaving a trace. Electronic records are not. A file can be copied endlessly, edited without a mark, and stripped of the very metadata that would show where it came from or whether it has been touched. Cybercrime evidence is especially fragile in this sense: server logs are routinely overwritten within days, cloud data may sit on machines outside the country investigating the offence, and a single record often passes through several different systems, each capable of altering it, before it ever reaches a courtroom. Courts everywhere have recognised that ordinary rules built for paper do not translate neatly to computer output, and each of the three jurisdictions in this paper responded to that recognition by legislating some threshold check before such evidence reaches the fact-finder. Where they differ sharply is in how strict that threshold check is, and in what happens when it is not perfectly met.

3 INDIA: THE CERTIFICATE AS ABSOLUTE GATEKEEPER

3.1 Section 65B and Its Troubled History

India's rule began with Section 65B of the Indian Evidence Act, 1872, inserted in 2000 alongside the Information Technology Act.[5] Section 65B(4) required a signed certificate, identifying the electronic record, describing how it was produced, and confirming the conditions of the computer that produced it, before any copy of an electronic record could be treated as proof of its contents.

The Supreme Court's early interpretation of this provision oscillated sharply. In State (NCT of Delhi) v. Navjot Sandhu, arising out of the 2001 Parliament attack case, the Court held that electronic records, including mobile telephone call records, could be proved through the ordinary rules on secondary evidence even without a Section 65B certificate.[6] That relaxed position lasted nearly a decade, and was

[4] Fed. R. Evid. 902(13)–(14).

[5] Indian Evidence Act, 1872, § 65B (inserted by Information Technology Act, 2000, § 92 & Sch. II, and repealed by the Bharatiya Sakshya Adhiniyam, 2023, with effect from July 1, 2024).

[6] State (NCT of Delhi) v. Navjot Sandhu, (2005) 11 SCC 600, 600–05.

followed as late as 2015 in Tomaso Bruno v. State of Uttar Pradesh, without reference to the intervening decision in Anvar.[7] In Anvar P.V. v. P.K. Basheer, a three-judge bench held that Section 65B is a complete and exhaustive code for electronic evidence, so the certificate is mandatory and Navjot Sandhu's contrary approach was wrong.[8] In 2018, a two-judge bench in Shafhi Mohammad v. State of Himachal Pradesh tried to soften this again, holding that the certificate requirement could be relaxed where the party seeking to rely on the record was not in control of the device that produced it.[9]

The pendulum finally stopped, at least for now, in Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal.[10] A three-judge bench reaffirmed Anvar and overruled Shafhi Mohammad, holding that the certificate is a condition precedent to admissibility, a requirement that cannot be dispensed with even where a party genuinely cannot obtain it. Tellingly, the facts of the case show exactly how this plays out in practice. Election-related compact discs had been produced without the required certificate, and both the trial court and the High Court found there was substantive compliance through oral evidence given in cross-examination. The Supreme Court held this was not enough: the certificate was indispensable regardless of what the oral evidence otherwise proved. The Court did soften the blow slightly by holding that a party can apply to the court for a direction compelling production of the certificate where the person or authority who should sign it refuses or fails to respond, and that the certificate could still be supplied at a later stage of the trial.[11] But the core rule survived intact: no certificate, no admission, whatever else the record might show.

3.2 The Certificate Gets Harder: Section 63 of the BSA, 2023

India replaced the Evidence Act with the Bharatiya Sakshya Adhiniyam, 2023, effective from 1 July 2024.[12] Section 63 is the direct successor to Section 65B and largely re-enacts the same scheme, but it goes further still. Section 63(4), read with the Schedule to the Adhiniyam, now requires a two-part certificate: Part A, filled in by the person in charge of the device, and Part B, filled in by an independent expert, disclosing the hash value of the record.[13] Where Section 65B needed one signature, Section 63 needs two, from two different people, one of whom must be a forensic expert.

This change did not go unchallenged. The Pune Bar Association petitioned the Supreme Court, arguing that Section 63(4) imposes undue hardship on ordinary litigants by requiring a hash-value disclosure and an expert's signature as a precondition to admissibility, and that this made the provision arbitrary and unconstitutional.[14] In 2026, the Supreme Court upheld the provision, reasoning that the risk of manipulation through artificial intelligence and deepfake technology gave the hash-value and expert-certification requirements a rational connection to the law's purpose.[15] Whatever the merits of that reasoning, the practical result is that India's certificate requirement has become more demanding, not less, even as the comparators examined in this paper moved in the opposite direction.

---

[7] Tomaso Bruno v. State of U.P., (2015) 7 SCC 178.
[8] Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473, 473–74.
[9] Shafhi Mohammad v. State of Himachal Pradesh, (2018) 2 SCC 801.
[10] Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1, 1–4.
[11] *Id.* at 4.
[12] Bharatiya Sakshya Adhiniyam, 2023 (brought into force July 1, 2024).
[13] *Id.* § 63(4) & Sch.
[14] Pune Bar Ass'n v. Union of India, Writ Petition (S. Ct. India 2026), *as discussed in SC Upholds Section 63(4) as Hash Value Requirement Ensures Authenticity of Electronic Evidence*, TAXGURU (May 28, 2026), https://taxguru.in/corporate-law/sc-upholds-section-634hash-requirement-ensures-authenticity-electronic-evidence.html.
[15] *Id.*

3.3 Why This Bites Especially Hard in Cybercrime Cases

Cybercrime prosecutions are where the certificate requirement causes the most damage, for three structural reasons. First, the person in charge of the relevant computer or communication device is often not a witness the investigating agency controls. It may be a foreign social media platform, a private telecom operator, or a bank's information-technology department, none of whom are eager to travel to an Indian trial court to sign a certificate. Second, cybercrime evidence typically passes through many hands: the victim's device, an internet service provider's server, a forensic laboratory, and the investigating officer, and Section 63(4) now asks for signatures from at least two of them in a specific format, raising the odds that at least one link in the chain is missing or defective. Third, investigating officers, who are usually not lawyers, may misunderstand or omit the certificate at the time of seizure, and by the time the omission is noticed, the original device or log may no longer be available to fix it.

4 THE UNITED KINGDOM: FROM CERTIFICATE TO JUDICIAL DISCRETION

The United Kingdom once had a rule strikingly similar to India's. Section 69 of the Police and Criminal Evidence Act 1984 provided that a statement in a document produced by a computer was inadmissible unless the party could show there were no reasonable grounds to believe the statement was inaccurate because of improper use of the computer, and that the computer was operating properly at the relevant time.[16]

Section 69 proved unworkable almost immediately. It produced technical litigation over trivial defects, most famously in Director of Public Prosecutions v. McKeown, where a defendant argued that a breathalyser's slow internal clock undermined the reliability of a reading, even though the clock error had nothing to do with the alcohol measurement itself. The House of Lords held that only malfunctions bearing on the accuracy of the document's contents mattered, not any and every defect in the machine.[17] The Law Commission examined the provision and, in a 1997 report, recommended its outright repeal, concluding among other things that the section did not address the real causes of unreliable computer evidence and that parties relying on such evidence were often in no position to satisfy a court about the internal operation of a machine they had not built and did not control.[18] Parliament agreed, and section 60 of the Youth Justice and Criminal Evidence Act 1999 abolished section 69 outright, with no replacement certificate scheme of any kind.[19]

Since 1999, computer evidence in the United Kingdom has been governed by an ordinary common law presumption: in the absence of evidence to the contrary, a court presumes that a mechanical or computer system was working properly at the relevant time.[20] If the opposing party puts forward some evidence that it may not have been working properly, the burden shifts to the party relying on the record to prove reliability. Any residual hearsay problem is handled through the ordinary hearsay provisions of the Criminal Justice Act 2003 rather than through any electronic-evidence-specific certificate.[21] The result

[16] Police and Criminal Evidence Act 1984, c. 60, § 69 (UK) (repealed 1999).
[17] Dir. of Pub. Prosecutions v. McKeown, [1997] 1 W.L.R. 295 (H.L.) (appeal taken from Eng.).
[18] LAW COMM'N, EVIDENCE IN CRIMINAL PROCEEDINGS: HEARSAY AND RELATED TOPICS, Law Com. No. 245, Cm. 3670 (1997) (UK).
[19] Youth Justice and Criminal Evidence Act 1999, c. 23, § 60 (UK).
[20] CROWN PROSECUTION SERVICE, LEGAL GUIDANCE: COMPUTER RECORDS EVIDENCE (UK) (setting out the common law presumption of proper functioning following the repeal of section 69).
[21] Criminal Justice Act 2003, c. 44, §§ 127, 129 (UK).

is that a genuine, reliable printout or log is not excluded merely because nobody signed a particular form. The record is treated like any other piece of evidence, admissible unless there is a real, evidenced reason to doubt it.

5 The United States: Many Roads to Authentication

The United States never adopted a certificate requirement at all. Federal Rule of Evidence 901(a) states the baseline: a proponent need only produce evidence sufficient to support a finding that the item is what it is claimed to be.[22] That standard can be met by live testimony from someone with knowledge, by the distinctive characteristics of the record itself, by expert comparison, or by several other means the rule expressly lists as illustrative rather than exhaustive.[23]

In 2017, the Federal Rules of Evidence were amended to add Rules 902(13) and 902(14), which allow electronic evidence to be self-authenticating, meaning admitted without live testimony, where a qualified person certifies that the record was generated by a reliable process, or that a copy was made using a verified method such as matching hash values.[24] Crucially, this certificate is optional. It is a convenience that spares a party the expense of calling a witness when authenticity is not seriously disputed, not a precondition without which the evidence cannot be admitted at all.[25] A party can always fall back on the ordinary, testimony-based routes under Rule 901 if no certificate is available.

The leading judicial explanation of how this plays out is Lorraine v. Markel American Insurance Co., where a federal magistrate judge set out, in detail, the full menu of ways a party can authenticate electronically stored information, stressing that counsel must be prepared to use whichever route the facts of the case actually support.[26] The flexibility of the American approach does not mean anything goes. Courts still exclude evidence when no adequate authentication is offered by any route. In United States v. Vayner, a printout of a social media profile page was excluded because the prosecution offered no evidence, beyond the page itself containing the defendant's name and photograph, that the defendant had created or controlled it, and the Second Circuit held that a reasonable juror could not find the page was what the government claimed.[27] The American system, in other words, filters out weak evidence through a searching, fact-specific inquiry into actual reliability, rather than through a single missing signature on a form.

6 Why India's Model Is the Least Workable of the Three

The comparison exposes what is really wrong with India's approach: it conflates a genuinely useful evidentiary safeguard with an inflexible, form-driven precondition. All three countries want the same thing: assurance that an electronic record actually is what it claims to be. The United Kingdom and the United States pursue that assurance through open-ended inquiries into actual reliability, a presumption that can be rebutted by real evidence in the former, and a menu of authentication routes tested against the facts in the latter. India instead asks a single, narrow, procedural question, namely whether a specific certificate,

[22] Fed. R. Evid. 901(a).
[23] *Id.* 901(b).
[24] *Id.* 902(13)–(14).
[25] Fed. R. Evid. 902(13)–(14) advisory committee's note to 2017 amendment.
[26] Lorraine v. Markel Am. Ins. Co., 241 F.R.D. 534, 538, 545–46 (D. Md. 2007).
[27] United States v. Vayner, 769 F.3d 125, 131–33 (2d Cir. 2014).

now in two parts, was properly executed, and makes the answer to that question dispositive, regardless of what other proof of reliability exists.

This produces exactly the outcome the Arjun Panditrao facts illustrate: a court can be entirely satisfied, from oral testimony given under cross-examination, that a record is genuine, and still be required to exclude it because a signature is missing.[28] In a cybercrime case, the record excluded this way is not a hypothetical. It might be the only chat log connecting an accused to a phishing scheme, or the only server log showing unauthorised access to a hospital's patient database. When that evidence is thrown out for a certification defect that has nothing to do with whether the log is genuine, the prosecution does not just lose a piece of evidence; it often loses the case, because electronic evidence is frequently all that exists in a cybercrime matter. The result is what this paper's title calls the certificate trap: a rule meant to keep out unreliable evidence ends up keeping out reliable evidence too, simply because reliability was demonstrated the wrong way.

The 2026 decision in Pune Bar Association v. Union of India forecloses a constitutional fix, since the Supreme Court has now held the two-signature version of the requirement to be a rational, permissible response to the risks of digital manipulation.[29] That may well be correct as a matter of constitutional law. Deepfakes and artificially generated media are real and growing threats to the authenticity of digital records, and a legislature is entitled to respond to them. But the fact that a rule survives a rationality challenge does not mean it is good policy. The United Kingdom and the United States show that it is possible to guard against manipulated evidence without making a single missing form fatal to an otherwise strong case. Reform, if it is to come, will have to come from Parliament rather than from the courts.

7 A WORKABLE PATH FORWARD

India does not need to give up on certification altogether. Hash values and expert sign-off are genuinely useful tools, especially given the real risk of manipulated digital media. What India should give up is treating the certificate as the only door into the courtroom. Three changes would bring India closer to the more workable models examined above, without abandoning the safeguards Parliament clearly still wants.

First, Section 63 should be amended so that the certificate operates as one recognised, convenient route to admission, as it does under Rules 902(13) and 902(14) in the United States, rather than as the exclusive route. Where a party can otherwise satisfy the court that a record is genuine, through witness testimony, forensic comparison, or independent corroboration, the absence of a certificate should go to weight, not admissibility.

Second, India could adopt something like the United Kingdom's rebuttable presumption for records generated by systems not reasonably in dispute, such as routine server logs and standard telecom call records, while reserving a stricter certification requirement for cases where the opposing party raises a genuine, evidenced objection to reliability. This would target the safeguard at cases that actually need it, instead of applying it uniformly regardless of whether anyone disputes the record's authenticity.

Third, the court-ordered production remedy recognised in Arjun Panditrao, currently limited to compelling an unwilling custodian to produce a certificate, should be broadened into a general judicial power

[28] Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1, 3–4.
[29] Pune Bar Ass'n v. Union of India, Writ Petition (Civil) No. 599 of 2026.

to accept alternative proof of authenticity, such as expert testimony on metadata or hash verification performed after the fact, whenever a certificate cannot reasonably be obtained, whether because a foreign platform will not cooperate or a custodian has died or disappeared. Investigating agencies should also be given clear, simple standard operating procedures, so that the two-part certificate is completed at the point of seizure rather than reconstructed, often unsuccessfully, months later at trial.

None of this requires India to copy the United Kingdom or the United States wholesale. It requires only that India stop treating a paperwork requirement as more important than the truth it was designed to protect.

8 Conclusion

Electronic evidence now decides most cybercrime cases before a single witness is cross-examined on the merits, because admissibility is judged first. India's certificate rule, tightened rather than loosened by the Bharatiya Sakshya Adhiniyam, 2023, and now upheld against constitutional challenge, continues to treat that certificate as an absolute gatekeeper: no certificate, no case, regardless of what else the record shows. The United Kingdom abandoned an almost identical rule in 1999 because it recognised that the rule did not track actual reliability. The United States never adopted such a rule at all, and instead built a flexible, multi-track system that still manages to exclude genuinely unauthenticated evidence when the facts warrant it. India can keep the parts of its certification scheme that guard against real risks like digital manipulation, while giving up the part that turns a missing signature into an automatic acquittal. Until it does, the certificate trap will keep swallowing genuine evidence, and with it, genuine prosecutions.

* * *

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.3

CHAPTER 3

When the Certificate Lies: Rethinking Authenticity Under the Bharatiya Sakshya Adhiniyam

RACHEL DELFIN*

The Bharatiya Sakshya Adhiniyam, 2023 replaced Section 65B of the Indian Evidence Act with Section 63. Section 63 keeps the certificate-based method that has controlled whether electronic records can be used in court since the cases Anvar P.V. v. P.K. Basheer (2014) and Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal (2020). The paper asks whether Section 63's continuity is a strength or a blind spot. Section 63's certificate is a two-part Schedule requiring a hash value and two signatures, designed to verify the chain of custody of a record. The certificate shows that a computer output accurately reproduces what a device captured. Section 63 assumes that an authentic event exists and that the record can be traced back to that event. AI-generated deepfakes break that presumption. A synthetic video can meet every requirement of Section 63. The synthetic video can carry the correct device details, the correct hash, and the correct signatures yet it can show an event that never happened. During 2025–26, High Court injunctions against deepfakes of public figures show that courts are already improvising remedies under Article 21 personality rights. The lack of guidance on how to authenticate content instead of just certifying its custodial trail explains this. The paper argues that the BSA's electronic evidence provisions modernize the language used for admissibility but do not solve the underlying epistemic problem, and also argues that the IT Rules of 2026 which require labeling of content cannot replace a real authentication standard.[1] Drawing on the text of the certificate Schedule, recent case law and a comparison with the UK Forensic Science Regulator's Code of Practice, the paper proposes a narrower reform. The reform would add a supplementary authenticity threshold for content that is plausibly generated or altered by AI. This threshold would be separate from the custodial certificate.

Keywords: Electronic evidence, Bharatiya Sakshya Adhiniyam, Deepfakes, Section 63, Authentication.

1 INTRODUCTION

Imagine scrolling through media and seeing a video of a public figure saying something controversial. The image is sharp, the person looks real, the voice sounds like them and their movements are natural. There's no sign that anything is wrong. In a world where artificial intelligence can create videos, voices and images, the problem isn't just whether a digital file was changed after it was made. The bigger challenge is whether the event in the video actually happened at all. This question becomes especially serious when such content ends up in a courtroom. Today electronic evidence plays a role in legal cases. Courts regularly use CCTV tapes, phone recordings, emails, photos, social media posts, call logs and other digital materials to help determine what really occurred. Because of this the law needs rules for deciding when such material can be accepted as proof. For a time Indian law dealt with this by focusing on authentication

* Law Student at Christ (Deemed to be University), Bengaluru, Karnataka, India.

[1] Ministry of Electronics and Information Technology, Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, notified Feb. 10, 2026 and effective Feb. 20, 2026, introducing mandatory labelling and provenance requirements for synthetically generated information.

and custody. If someone could show how an electronic record was created, stored, copied and handled properly and meet the conditions set out in the law, courts would accept it as reliable.

A key part of this process came from Section 65B of the Indian Evidence Act, 1872. Over time Supreme Court rulings in cases like State (NCT of Delhi) v. Navjot Sandhu, Anvar P.V. v. P.K. Basheer and Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal helped shape how electronic evidence is admitted. The Anvar case made it clear that following Section 65B is essential to allow records into court—especially if they're presented in the way the law allows. The Arjun Panditrao decision reinforced the need for certification and clarified that the certificate is unnecessary where the original electronic record itself is produced. Now the Indian Evidence Act has been replaced by the Bharatiya Sakshya Adhiniyam, 2023. Section 63 of the law covers the admissibility of electronic records[2] and keeps much of the old structure from Section 65B. It includes a Schedule that requires information about the record itself, the device or source it came from and its hash value. This Schedule gives an organized way to verify the origin and integrity of digital files. But here's the issue: deepfakes break this system. A deepfake may be an authentic digital file, not edited after creation, correctly copied and verified through a hash value. The file could have come from a device, passed through proper channels and even have a valid certificate. Still the event it shows might never have taken place. A falsehood, even if preserved exactly as it was, is still a falsehood. That's where the current legal framework falls short.

This distinction forms the central argument of this paper. The existing certificate-based approach is primarily concerned with the authenticity and integrity of the electronic record as an object of evidence. Deepfakes introduce a different question concerning the authenticity of the reality represented by that record. Therefore, the issue is not whether certification is unnecessary. Rather, the issue is whether certification alone can remain sufficient when the real dispute concerns the truthfulness of the underlying event. The central research question of this paper is therefore: How far can a custody-based certificate regime remain adequate once the authenticity of the underlying event, not merely its custodial trail, is what is actually in dispute? This paper first examines the development of the law relating to electronic evidence under Section 65B of the Indian Evidence Act. It then considers the continuity between Section 65B and Section 63 of the BSA, particularly the importance of certification, device information and hash values. The paper then examines the challenge posed by deepfakes and distinguishes between what may be termed "custodial authenticity" and "substantive authenticity." Finally, it proposes a framework called C.A.S.E., or the Certification–Authenticity Split for Electronic Evidence, under which statutory certification remains the ordinary first layer of authentication, while a specific and credible challenge to the authenticity of the underlying content can trigger a second layer of substantive verification. The objective is not to argue that every electronic record should automatically be subjected to forensic examination. Such an approach would be impractical and could place an unnecessary burden on courts, litigants and already limited forensic resources. Instead, the proposed approach attempts to create a middle ground between complete reliance on certification and compulsory forensic verification of every digital record.

2 THE OLD REGIME COMPRISING SECTION 65B AND ITS CASE LAW

The legal treatment of electronic evidence in India developed in response to a basic difficulty. Traditional evidence law was largely designed around physical documents, while digital information could exist

[2] Bharatiya Sakshya Adhiniyam, 2023, § 63 and Schedule.

simultaneously in several forms and could be copied without an obvious physical distinction between an original and a duplicate. A photograph stored on a phone, for example, could be transferred to a computer, uploaded to a cloud service and reproduced on another device without the ordinary physical characteristics associated with conventional documents. The introduction of Sections 65A and 65B into the Indian Evidence Act tried to solve a problem by setting up a way to prove the contents of electronic records. At first the Supreme Court took a flexible view in the case of State (NCT of Delhi) v. Navjot Sandhu.[3] In that case it looked at whether electronic records could be admitted as evidence using the existing rules for evidence. Later the Supreme Court changed its position in the case of Anvar P.V. v. P.K. Basheer.[4] In this case the court treated Section 65B as the rule that controls how electronic evidence is accepted when the situation fits what the section describes. The court made it clear that any electronic evidence given in the way laid out by Section 65B must meet legal requirements. One of those requirements is the certificate mentioned in Section 65B(4). This decision was significant because it shifted the law away from treating records simply as another type of written document. Instead it acknowledged that electronic records need their own set of rules and protections. The value of the certificate becomes clear when we look at what Section 65B was meant to do. The certificate gives details about the record itself, how it was created and which computer or device was used. It acts like a kind of proof that shows how the record came to exist and how it was produced before being presented in court. This helps ensure that the record is reliable and trustworthy.

This approach was subsequently considered in detail in *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal*. The Supreme Court reaffirmed the position in *Anvar* and treated the Section 65B certificate as an important requirement for the admissibility of electronic evidence falling within the provision. At the same time, the Court clarified that where the original electronic record itself was produced as primary evidence, the situation was different from the production of a computer output or secondary electronic evidence.[5]

The decision also recognised a practical difficulty. A party seeking to produce electronic evidence may not always have control over the device or computer from which the record originated. The Court therefore acknowledged that a court could assist in obtaining the necessary certificate in appropriate circumstances rather than allowing the absence of a certificate, where the party genuinely could not obtain one, to automatically defeat the claim. The development from *Navjot Sandhu* through *Anvar* to *Arjun Panditrao* demonstrates an important movement in Indian evidence law. The law became increasingly concerned with the reliability of the process through which electronic information was produced and brought before the court. The certificate was therefore not merely a technical formality. It functioned as a legal mechanism through which the court could be given assurance about the electronic record and the system that produced it. However, this model rests on an assumption that becomes increasingly complicated in the age of generative artificial intelligence. The certificate can provide evidence that a particular digital file came from a particular device or source and that the relevant statutory conditions were satisfied. It can therefore help answer the question: "Is this the same electronic record that was produced and preserved?" It does not necessarily answer a different question: "Does the event represented by this

[3] State (NCT of Delhi) v. Navjot Sandhu, (2005) 11 SCC 600.
[4] Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473.
[5] Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1.

electronic record actually correspond to something that happened in the real world?" This distinction becomes crucial when the record itself is capable of being artificially generated.

3 CHANGES WHICH OCCURRED UNDER SECTION 63 BSA

The BSA attempts to modernise the law of evidence and specifically recognises electronic and digital records. Section 61 recognises electronic or digital records, while Section 63 deals with the admissibility of electronic records.[6] In this sense, the BSA acknowledges the increasing importance of digital information within the evidentiary process. Section 63 substantially continues the logic of Section 65B. It establishes conditions relating to the computer or communication device, the regular use of the device, the ordinary course of activities and the production of the electronic record. Most importantly for the present discussion, Section 63(4) is accompanied by a Schedule containing a prescribed certificate.

The Schedule requires the party producing the electronic record to provide details regarding the source of the record. These may include a computer, storage media, DVR, mobile phone, flash drive, CD/DVD, server or cloud. It also asks for identifying information relating to the device, such as the make and model, serial number and relevant identifiers. The certificate further addresses matters such as lawful control of the device, whether it was functioning properly and whether information was regularly fed into the device in the ordinary course of activities.[7]

One of the most notable aspects of the Schedule is its treatment of hash values. A hash may be understood, at a basic level, as a kind of digital fingerprint for a file. When a particular electronic file is processed through a hashing algorithm, it produces a corresponding value. If the file is subsequently changed, even in a small way, the resulting hash may change. The Schedule therefore provides a mechanism for recording the hash value of the electronic or digital record and identifying the algorithm used to generate it. This is an important development because it strengthens the ability of the legal system to establish the integrity of an electronic file. However, it is important not to give the hash function a role that it cannot perform. A hash can assist in establishing that a particular file has remained unchanged from the point at which its hash was calculated. It cannot establish that the contents of the file depict a genuine event. This can be illustrated through a simple example. Suppose a person creates a completely fabricated video using artificial intelligence and saves it on a mobile phone. The video is then copied from the phone and its hash value is calculated. If the file is later produced in court and the hash value matches the certified value, the integrity of that particular file may be established. But the hash does not establish that the person shown in the video actually said or did what the video depicts. In other words, the hash can demonstrate file integrity, but file integrity is not the same as event authenticity. This distinction suggests that Section 63 should be understood as an important procedural development rather than a complete conceptual transformation of electronic evidence law. The BSA provides more structured information regarding the device, the source and the integrity of the record. These safeguards are valuable. Yet they primarily address the history and integrity of the electronic record rather than the truthfulness of the reality represented by the record. This is where deepfakes create a particularly difficult evidentiary problem.

[6] Bharatiya Sakshya Adhiniyam, 2023, §§ 61–63.
[7] *Id.* Schedule to § 63.

4 WHERE THE REGIME BREAKS THROUGH THE DEEPFAKE ERA

A deepfake is synthetic or manipulated media created using artificial intelligence or other digital techniques to make a person appear to say or do something that they did not actually say or do. The technology itself is not limited to videos. Artificial intelligence can also be used to generate or manipulate photographs, audio recordings and other forms of digital content. The evidentiary difficulty created by deepfakes is fundamentally different from ordinary questions of tampering. Traditional digital evidence analysis often asks whether a file has been modified after its creation. Deepfakes complicate this assumption because the file may have been created as a manipulated representation from the beginning. Consider a hypothetical case involving a video allegedly showing a company director approving an unlawful transaction. The video is stored on a mobile phone and produced before a court. The party producing it obtains the necessary certificate under Section 63. The phone was functioning properly, the file was stored on the device and the hash value corresponds with the copy presented to the court. From the perspective of the statutory requirements, the record may appear properly authenticated. Yet suppose the video was generated entirely through artificial intelligence. The director never attended the meeting and never made the statement contained in the video. The problem is therefore not that someone altered the file after it was created. The problem is that the file itself was created to represent an event that never occurred.

This creates a distinction between two different forms of authenticity. The first may be described as custodial authenticity. It concerns whether the electronic record presented to the court is the same record that was obtained from the relevant device or source and whether its integrity has been maintained. Section 63 is particularly useful for this purpose. The device details, certification requirements and hash value can all contribute to establishing this form of authenticity. The second may be described as substantive authenticity. This concerns whether the content of the electronic record accurately represents the underlying event, person, statement or conduct that it purports to show. Substantive authenticity asks whether the person actually spoke those words, whether the event actually occurred and whether the recording is a genuine representation of reality.

The two forms of authenticity may overlap, but they are not identical. A court could therefore encounter a situation in which an electronic record is completely authentic as a digital object but completely false as a representation of an event. This is the central weakness that deepfakes expose in a purely custody-based approach. Recent Indian litigation demonstrates that courts are increasingly being confronted with the consequences of AI-generated and manipulated content. In *Shilpa Shetty Kundra v. Getoutlive.in & Ors.*, the Bombay High Court dealt with material involving AI-generated deepfake content and considered issues relating to privacy, dignity and personality interests.[8] Although such proceedings do not themselves resolve the evidentiary question posed in this paper, they demonstrate the increasing legal significance of synthetic media. The growing use of deepfakes also means that courts cannot simply assume that a visually convincing recording is reliable because it appears to have originated from a genuine device. At the same time, it would be equally problematic to assume that every digital record is potentially fabricated and therefore requires forensic examination.

[8] Shilpa Shetty Kundra v. Getoutlive.in (Bom. HC Dec. 26, 2025). *See also*, on AI-generated content and personality rights, Karan Johar v. Indiawaale.com (Del. HC Sept. 19, 2025); Aishwarya Rai Bachchan v. Vroniq Consulting Pvt. Ltd. (Del. HC Sept. 9, 2025); Hrithik Roshan v. Ashok Kumar (John Doe), CS(COMM) No. 1107 of 2025 (Del. HC).

This creates the need for a balanced framework. A blanket requirement that every video, photograph, audio recording or electronic document be subjected to forensic examination before being relied upon would be difficult to implement. It would increase litigation costs, delay proceedings and place considerable pressure on forensic laboratories and experts. It could also create an unnecessary barrier to the admission of ordinary electronic evidence where there is no genuine reason to doubt its authenticity. The better approach is therefore to retain certification as the ordinary first stage while recognising that certain circumstances should trigger additional scrutiny.

5 THE C.A.S.E. FRAMEWORK (CERTIFICATION–AUTHENTICITY SPLIT FOR ELECTRONIC EVIDENCE)

The central proposal of this paper is the C.A.S.E. Framework, or the Certification–Authenticity Split for Electronic Evidence. The framework is based on a simple proposition: certification and substantive authenticity should not be treated as if they answer the same question. The first stage of the framework is Certification. Section 63 should continue to operate as the ordinary starting point for electronic evidence. The certificate and the accompanying information regarding the source device, manner of production, working condition and hash value perform an important function. They help the court determine whether the electronic record has been properly identified and whether its integrity has been preserved. There is no need to discard this mechanism merely because deepfakes exist. In fact, doing so would create unnecessary uncertainty in the law of electronic evidence. The certificate remains valuable because an electronic record must first be identified and its custodial integrity established before questions concerning its substantive authenticity can meaningfully arise.

The second stage is Authenticity Challenge. Additional scrutiny should arise only when the opposing party raises a specific and reasonably credible challenge to the authenticity of the underlying content. The mere possibility that artificial intelligence could have been used should not be sufficient. If that were the rule, practically every digital recording could be challenged on the basis that it might be synthetic. Instead, the challenge should contain some identifiable basis for suspicion. For example, a party may point to unusual inconsistencies in the audio or video, unexplained changes in lighting or facial movement, inconsistencies between the metadata and the alleged circumstances, evidence that the person shown was elsewhere at the relevant time, discrepancies between the recording and independently established facts, or other circumstances suggesting manipulation.

The purpose of this stage is not to require the challenging party to prove the deepfake immediately. That would place an unrealistic burden on the party who is precisely seeking access to verification. The purpose is to establish a threshold at which the ordinary certification process is no longer sufficient to resolve the dispute. The third stage is Substantive Verification. Once a credible authenticity challenge has been raised, the court may require additional evidence concerning the substantive authenticity of the record. This could include forensic examination of the audio or video, analysis of metadata, examination of the source device, frame-level analysis, examination of encoding patterns, comparison with verified recordings of the relevant person, or other appropriate technical methods. Importantly, this stage should not assume that one particular AI-detection technology is always capable of determining whether a recording is genuine. Detection technologies themselves are developing and may produce inaccurate results.

Therefore, substantive verification should ideally be based on a combination of technical and contextual evidence rather than a single automated conclusion.

For example, if a video allegedly shows a person speaking at a particular conference, forensic examination might be combined with evidence concerning whether the person was actually present at the venue, recordings from other cameras, photographs taken at the event, witness testimony and original device information. The question before the court would then become broader than simply whether a file had been altered. The final stage is Evidentiary Responsibility. Once a credible challenge has crossed the threshold for substantive verification, the party relying upon the electronic record should ordinarily be expected to provide reasonable additional support for its authenticity, particularly where that party has greater access to the original device, source material or surrounding evidence. This should not be understood as an automatic reversal of the burden of proof in every case. Nor should the existence of an authenticity challenge automatically make the electronic record inadmissible. Instead, the court should consider the strength of the challenge, the availability of the relevant evidence and the circumstances of the case when deciding how much additional proof is necessary. The significance of the C.A.S.E. framework therefore lies in its separation of two questions that are often treated as though they were identical. The first question is whether the electronic record is genuine as a record. The second is whether the content of that record genuinely represents the event it claims to represent.

The first question can ordinarily be addressed through Section 63. The second may require additional evidence when a credible challenge is raised. Such an approach also has the advantage of being more realistic for the Indian legal system. India's forensic infrastructure has important capacity constraints, and requiring forensic examination of every electronic record would not be an efficient use of limited resources. A trigger-based system would allow forensic expertise to be directed towards cases where the risk of manipulation is actually material. The approach may also be gradually implemented. Courts could initially apply the framework in cases involving high-risk digital evidence, such as disputed videos, audio recordings or images where the authenticity of the underlying event is central to the dispute. Over time, clearer judicial standards could develop concerning what constitutes a sufficient authenticity challenge and what types of verification should ordinarily follow.

The United Kingdom provides a useful comparative perspective in this regard. Its forensic evidence framework places considerable emphasis on standards, competence and quality management through the statutory role of the Forensic Science Regulator and the applicable Code of Practice.[9] The UK approach is not specifically designed as a deepfake rule and cannot simply be transplanted into Indian evidence law. However, it demonstrates the broader principle that forensic evidence requires institutional standards and reliable processes rather than merely the existence of technical expertise. For India, the important lesson is therefore not that every electronic record must undergo forensic testing. It is that when forensic examination becomes necessary, the examination itself should be carried out according to reliable standards and by appropriately qualified experts. The C.A.S.E. framework consequently seeks to preserve the advantages of the existing statutory system while responding to the new epistemic problem created by synthetic media. It does not reject certification; it places certification in its proper evidentiary role.

[9] FORENSIC SCIENCE REGULATOR, CODE OF PRACTICE (VERSION 2), issued pursuant to the Forensic Science Regulator Act 2021.

6 CONCLUSION

The development of Indian electronic evidence law shows a clear attempt to adapt traditional evidentiary principles to a digital environment. Section 65B of the Indian Evidence Act created a specialised mechanism for electronic records, and the Supreme Court's decisions in *Anvar P.V.* and *Arjun Panditrao* provided important guidance concerning certification and admissibility. The BSA has continued this approach through Section 63 while making the certification process more detailed, including through the Schedule's requirements concerning device information and hash values. These developments are important, but the rise of deepfakes reveals a conceptual limitation that cannot be solved merely by making the certificate more detailed. The certificate can establish important facts about the electronic record. It can help establish where the record came from, how it was produced and whether the particular file has remained unchanged. A hash value can strengthen the proof of digital integrity. However, none of these mechanisms necessarily establish that the event represented by the recording actually occurred. This distinction is increasingly significant because artificial intelligence has changed the nature of digital manipulation. Earlier, the central evidentiary concern could often be expressed as: "Has this recording been altered?" With deepfakes, the more fundamental question may be: "Was there ever a real event corresponding to this recording?"

The answer requires a shift from treating authentication as a single concept to recognising different layers of authenticity. The proposed C.A.S.E. framework attempts to provide such a distinction. Certification remains the ordinary first layer because Section 63 serves an important purpose in establishing the integrity and provenance of electronic evidence. However, where a specific and credible challenge is raised regarding the authenticity of the underlying content, the court should have the ability to move to a second layer involving substantive verification. This may involve forensic analysis as well as independent corroborative evidence. The level of scrutiny should depend upon the circumstances of the case rather than being imposed uniformly upon every electronic record. Such a model also avoids two extremes. On one side is complete reliance on certification, which risks treating a properly preserved falsehood as trustworthy merely because its digital chain of custody is intact. On the other side is universal forensic scrutiny, which would be impractical and could unnecessarily burden the judicial and forensic systems. The deeper lesson is that technological change does not always require completely new legal rules. Sometimes it requires existing legal concepts to be examined more carefully. Section 63 answers an important question about electronic evidence, but deepfakes demonstrate that it does not answer every question that a court may need to ask. Ultimately, the difference can be expressed through two simple questions: "Is this the same file?" and "Is what this file shows actually real?" Section 63 is well equipped to assist with the first question. The challenge for the future of Indian evidence law is to ensure that the second question does not disappear merely because the first one has been answered. The C.A.S.E. framework therefore proposes not a rejection of the existing law, but its development. In an environment where digital evidence can be preserved perfectly while depicting an event that never happened, custody should remain evidence of integrity, but it should not automatically become evidence of truth.

* * *

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.4

CHAPTER 4

When Seeing Is No Longer Believing: Deepfakes, Synthetic Evidence and the Criminal Process in India

SHIRVI GOYAL*

Deepfakes and similar technologies of synthetic media can generate such accurate images, videos, and audio, as to convincingly replicate real people and events. Although the technology might be utilized in crimes, such as fraud, identity theft, and image-based abuses, it poses a significant challenge to the role of digital evidence. The present paper discusses the likely and emerging challenges which deepfakes and synthetic media may present to criminal process in India. It interrogates whether the present law governing electronic evidence, including under the recently enacted Bharatiya Sakshya Adhiniyam, 2023, provides sufficient framework to deal with the issues of integrity, authenticity, and reliability that emerge from AI-generated or synthesized material. It argues that standard methods for electronic evidence may fail if issues related to the veracity of the recorded content are in conflict with the reliability of its preservation. Finally, the study addresses the role of forensic analysis, metadata and digital provenance in differentiation of true recordings from the synthesized. It evaluates the risk of the "liar's dividend" where authentic evidence can be disregarded because of readily available manipulation technologies. The research will conclude by recommending greater procedural and evidentiary protections that will allow criminal justice process to react to synthetic evidence without discrediting authentic digital recordings.

Keywords: Deepfakes; Synthetic Evidence; Digital Evidence; Criminal Process; Electronic Evidence.

1 INTRODUCTION

It is difficult not to assume that digital evidence will form a central feature of the criminal justice system moving forward. Records that may once have been overlooked by authorities, such as CCTV footage, mobile-phone records, photographs and messages are now routinely used as part of investigations and to prove charges in courts. These concerns were formally recognized by the Bharatiya Sakshya Adhiniyam, 2023, hereinafter the "BSA", giving these electronic and digital records the same status in law as documentary evidence, subject to their conditions of admissibility being met.[1]

This raises a new issue of the challenges presented by 'deepfakes'. Unlike standard digital records, a video may not actually reflect the event, or an image may not have been faithfully captured. They raise evidentiary problems in that such synthetic material may contain any statement made by anyone, and could result in evidence being wrongly taken that the defendant made or did whatever the relevant communication alleges. In many cases this is significant only in so far as it goes to establishing the individual identified as the defendant and what conduct they are alleged to have undertaken.

Indian courts, including the Supreme Court, have always treated electronic evidence very differently. In 2020, the Supreme Court re-affirmed many of these principles by emphasizing compliance with the strict statutory rules for proving electronic records as enshrined in the (then) Indian Evidence Act in a

* Student at Law College Dehradun, Uttaranchal University, Dehradun, Uttarakhand, India.

[1] Bharatiya Sakshya Adhiniyam, 2023, §§ 61–63.

landmark case called *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal*.[2] Whilst a new act has been put into place called the BSA it has retained specific parts relating to the authenticity, and admissibility of digital evidence. This compliance with technical regulations does not, however, make the content genuine and could result in the introduction to court of manipulated or synthetic content even though it does, under specific provision of the BSA fit the rules relating to admissibility of digital evidence as previously provided for.[3]

Criminals have always sought to perpetrate fraud against others through some method of manipulating real evidence in order to get persons to comply with whatever demands are made. While this is now achieved through new technologies, deepfakes create a two-prong attack, against individuals through both use in fraud or image-based abuse, as well as against the idea of individual guilt. An attack on both can make existing evidence seem more realistic, while in fact causing the opposite: where courts, witnesses or evidence purveyors might begin dismissing genuine audio-visual recordings as synthetic based on the mere fact that technically manipulation is possible – 'the liar's dividend.'[4]

This paper discusses the impact of deepfakes on the Indian criminal process. It explores the existing framework relating to digital evidence and focuses upon the evidentiary challenges of authenticating the integrity of synthetic data in investigations and in court. It is argued that while admissibility regulations have a role, the primary challenges lie in the forensic ability to prove what evidence is and is not synthetically generated.

2 DEEPFAKES AND THE EMERGING LANDSCAPE OF CYBERCRIME

2.1 Nature and Uses of Deepfake Technology

The deepfake represents a category of synthetic media that use an AI's manipulation or addition of visual or sonic components to someone's image or voice to generate an audiovisual recording where such a person appears to say or do something that didn't in fact happen. The issue the technology creates involves not just the manipulation of digital records, but the very increasing ability of such technology to produce an event that seems completely true to a user without expertise. Such technology has transformed from being exclusive to technically capable people to publicly available online.[5]

2.2 Deepfakes and Emerging Forms of Cybercrime

Criminal activity using deepfakes follows the spread of this accessibility. One application has already arisen under fraud and impersonation. Synthetic audio or video material may attempt to replicate a real person to imitate a communication and create the effect of reality. Such an event may prove advantageous within social engineering, directed fraud, impersonation-based deception, and other forms of financial cybercrime. An instance the FBI reported where this occurred included use within personalized social engineering schemes, spear-phishing attacks, business email compromise and similar fraud actions.[6] Such an advantage would become considerably amplified where a participant in a criminal transaction views

[2] Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1.
[3] Bharatiya Sakshya Adhiniyam, §§ 61–63.
[4] Bobby Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1785–86 (2019).
[5] FED. BUREAU OF INVESTIGATION, OVERSIGHT OF THE FBI CYBER DIVISION (Mar. 29, 2022).
[6] *Id.*

realistic visual or acoustic representations that seem to be evidence familiar to them from past occurrences, inducing trust without further evaluation.

Serious risk has also been developed for sexual assault, harassment and sextortion. A person's photographs or video taken earlier in a relationship or publicly available online can be altered to seem like the victim engaged in a sexual act. Subsequently the individual will be threatened with the publication of such an event, or be forced into providing cash or real sexual encounters. There are instances where minors, as well as adults, have become targets of deepfake images to facilitate sexual violence, extortion or other criminal acts.[7] Such consequences can, without true sexual content, be brought about by mere images depicting it.

Another area of concern is false attribution and identities. A synthetic performance might appear to record an individual at any specific time saying they did something that they hadn't, or being at a specific place when he or she wasn't. This is especially problematic when an individual may view fake material prior to evidence collection efforts being in place. Once distributed, fake images may impact many in the public sphere before their authenticity is verified.

2.3 The Liar's Dividend and Threats to Digital Evidence

Deepfakes present a problem in that people aren't just viewing evidence that is manipulated; they might find their own future use of a real image challenging where proof of that authenticity isn't easily established. This represents something more than another avenue for fraud – there is a potential challenge with verifying every digital piece of data for authenticity.[8] The FBI acknowledges this problem has and will lead to questions surrounding how to establish authenticity when looking at real digital evidence. The risk of deepfakes becomes even greater when looking at their role in the evidentiary process. An unauthentic performance of a crime may at first view be assumed as genuine traces leading an investigation. On the other hand, parties might attempt to challenge authentic visuals or acoustic performances as deepfakes, thereby introducing questions of how real footage ought to be used as proof. These issues need to be anticipated and considered as these technologies gain access throughout India. An increased access will force Indian judicial proceedings to take account of deepfake technology, thereby allowing for either their use within the fraudulent context, or demanding stricter standards for evidence.

3 Digital Evidence within the Indian Criminal Justice Landscape

3.1 Legal Framework for Electronic and Digital Evidence

The Indian criminal justice system has, to some extent, moved beyond the traditional notion that evidence is primarily composed of physical documents and oral testimony. Electronic evidence has become an inherent part of criminal investigation and trial, owing to the ubiquitous nature of mobile phones, computers, cloud storage, and digital communication systems. The Bharatiya Sakshya Adhiniyam, 2023 ("BSA"), which replaced the Indian Evidence Act, 1872, unequivocally brings electronic and digital

[7] Fed. Bureau of Investigation, Malicious Actors Manipulating Photos and Videos to Create Explicit Content and Sextortion Schemes, Public Service Announcement No. I-060523-PSA (June 5, 2023).
[8] Fed. Bureau of Investigation, Oversight of the FBI Cyber Division.

records under the ambit of evidence law. In particular, the BSA sections relating to electronic or digital records (sections 61-63) address their existence and admissibility.[9]

Section 61 explicitly provides that an electronic or digital record cannot be denied the quality of legal significance solely on the grounds that it exists in an electronic or digital form. Section 62 lays down the conditions for proving such a record, and section 63 defines their admissibility.[10] The legislative effort is, thus, a recognition of the reality that significant information can manifest and be preserved only in digital format. The BSA also makes it mandatory, for specific electronic records, to produce a certificate in the prescribed form which must include particulars about the device or source from which the record was derived and should attest to the hash value of the electronic or digital record.[11] The hash is valuable because it could assist in establishing whether the digital file produced later in time in court is the same as the digital file that was captured at the outset.

This statutory regime builds upon principles that were established by the Indian Supreme Court, prior to the enactment of the BSA, based upon the provisions of the Indian Evidence Act, 1872. In the case of *Anvar P.V. v. P.K. Basheer*, the Supreme Court had deemed the statutory requirements governing electronic evidence as a separate evidentiary regime.[12] This assertion was later clarified in the three-judge Bench judgment in *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal*, where it stated the significance of the certificate requisition, especially in the case of secondary electronic evidence, whilst distinguishing between the production of original electronic record versus submission in the output/copy produced in respect of the same.[13] It is therefore evident that the transition from section 65B of the former Act to section 63 of the BSA represents a significant continuity in our approach and a legislative streamlining.

Similarly, our procedural framework has seen increasing accommodation of digital modes of investigation. For instance, the Bharatiya Nagarik Suraksha Sanhita, 2023 ("BNSS") requires searches and seizures to be documented by means of audio-video recording by virtue of section 105.[14] This is an important procedural step as the manner in which devices or articles containing electronic information were discovered, seized and handled can impact the credibility of the evidence introduced at the trial.

3.2 Admissibility, Authenticity and Integrity

However, the existence of a statutory mechanism for admitting electronic records into evidence doesn't solve the challenge that deepfakes present. Admissibility and authenticity are two different, though related, issues. A recording may fulfil the procedure pertaining to its production and certification by the authorities but the contents within the said recorded materials could have been digitally manipulated to the core. While a certificate can vouch for the contents relating to how the record came into existence and from what source, it cannot, on its own, certify that every image, sound or utterance within such a record represents a real event. This becomes particularly pertinent when AI is used for altering the constituent components of the recorded materials.

Therefore, although the recognition of hash values and device-specific data in the BSA is of considerable help, it should not be conflated as a panacea to the synthetic data problem. A hash can attest to

[9] Bharatiya Sakshya Adhiniyam, §§ 61–63.
[10] *Id.* §§ 61–63.
[11] *Id.* § 63(4) & sched. (prescribing the certificate and requiring identification of the digital source and hash value).
[12] Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473.
[13] Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1.
[14] Bharatiya Nagarik Suraksha Sanhita, 2023, § 105.

the integrity of a file at a certain time of its retrieval but not necessarily to the fact that the file was not generated or manipulated prior to its capture. Deepfakes, hence, take the examination from the stage of integrity of a digital file to the question of how and from what event the file originated.

3.3 Limitations of the Existing Framework

This legislative framework has laid a foundation for dealing with electronic evidence in India, but deepfakes highlight a critical gap. The legal frameworks for electronic evidence were primarily aimed at the acknowledgment of, and procedures for, production and verification of electronic records. Synthetic media adds a new level: content-level authenticity. A determination as to whether a digital record is technically sound cannot always be equated with a determination as to whether it truthfully depicts reality, and herein lies the crux of the evidentiary difficulties posed by deepfakes.

4 Deepfakes as Synthetic Evidence: Problems with Authenticity and Reliability

4.1 Authenticity of Synthetic Evidence

The foremost evidentiary challenge posed by deepfakes is that an electronic file may be accurate in form, while being falsified in its representation of events. Existing rules regarding electronic evidence predominantly centre on establishing the identity of an electronic record and how it was created and preserved. Deepfakes add a separate issue: whether the content itself correctly depicts an event as it unfolded. The importance of this distinction can best be illustrated by an unaltered recording that portrays a fabricated event.

Under the BSA, a compliant electronic record, meaning, it satisfies all relevant provisions pertaining to it, will be admitted before a court. But the admission of this electronic record to be a record that was actually created, does not signify its content's truthfulness. For example, a photographic, audio, or video record, when properly taken from an electronic device, and its contents' integrity is maintained once recovered/obtained, will in truth not convey any meaningful information about an event if its contents themselves are a falsehood.[15]

At present, there are two levels of authentication involved. The first is authentication relating to the integrity of the record: where the produced electronic file is indeed the file collected or retained from the device. The second authentication relates to the authenticity of the content in the record: is the person in an audio/video recording actually saying what the record purports the person to say, did the event take place at all and is the person who appears to be speaking/participant truly the person whose statement is reported. This second level of authentication becomes acutely challenging with deepfakes.

4.2 Metadata, Hash Values and Forensic Examination

While metadata, time-stamps and hash values serve vital functions in the examination of an electronic document, they should not be conclusive proofs of a substantively true record. A hash value confirms that an electronic file remains unchanged from the time its hash value was computed; it cannot provide proof that the file was not modified prior to the computation of the hash value. Correspondingly,

[15] Bharatiya Sakshya Adhiniyam, §§ 61–63.

metadata only confirms aspects like the date a file was created, edited, or sent; however, the metadata may also be incomplete, edited or deleted with ease.

Technical methods of preservation are thus just one level of an examination, rather than a complete test for truth. The hash-value and certificate provisions of the BSA only enhance the chain of custody of an electronic document but cannot, in and of themselves, provide substantive proof of a seemingly factual deepfake video.[16]

Thus, the relevance of expert witness testimony is now heightened more than ever in deepfake cases. Authentication tests require analysis of audio characteristics, visual inconsistencies of facial structure/movement, file compression parameters, incorrect lighting as well as frame-level inconsistencies, etc., and only specialized practitioners equipped with sophisticated software can identify them. A major practical issue in such tests is the rapid evolution of creation technology, posing a risk to the reliability of existing detection techniques that identify certain markers of fabrication.

Indian evidence law already accepts witness expertise on 'questions as to identity of handwriting or finger impressions' as science or specialized knowledge wherein a court must form an opinion.[17] Likewise, expert analysis will have a role in the trial of deepfake evidence as it can guide the court's opinion on authenticity. But just as with any other evidence, expert testimony does not command or replace judicial assessment; the trial court will have to evaluate the entire evidence while keeping in mind the specialized nature of such proof and the methodology and materials involved in the forensic examination.

4.3 The Liar's Dividend and Evidentiary Reliability

The challenge operates in reverse also. Given how commonplace convincing fabrication may become in the future, a party could possibly claim that an authentic electronic recording depicting inconvenient actions on his part is a forgery without substantial technical grounds. This has the unfortunate consequence of creating what has come to be called "the liar's dividend": The higher the belief society develops of truly deceiving falsified records becoming increasingly easy to produce and use effectively, the easier it is for a fraudulent individual to deny or dismiss any genuine evidence of his actions.[18]

This creates a dilemma in the criminal justice process: The court can neither accept a digitally manufactured recording purely based on an external, convincing appearance; nor can a seemingly credible piece of evidence be rejected simply by virtue of the fact that scientific technology allows for alteration. Blindly trusting digitally recorded evidence, believing it to be as valid as the original, and deeply distrusting it, believing it to be fake, would both represent extremes and would be equally detrimental for the search for an authentically produced digital record. Unfortunately, this is reflected in the existing framework of proof.

While there are procedures governing electronic evidence, some involving the requirement that evidence must be verified (such as the maintenance of integrity of the digital evidence recorded), they do not deal specifically with deepfake-authenticity, and do not set standards for a judge on when to request forensic proof, to which level the data should be scrutinised, and how conflicting analyses on a deepfake's origin should be treated in the court.

---

[16] *Id.* § 63(4) & sched.
[17] *Id.* § 39.
[18] Chesney & Citron, *supra* note 4, at 1785–86.

That's the problematic gap: When this gap arises in the context of a criminal procedure, where incorrect use of fabricated electronic evidence can cause a catastrophe just as well as authentic evidence can be unjustly excluded, the ideal solution therefore consists in establishing clear rules that allow for better identification of legitimate as against expertly manufactured synthetics in order to preserve a justice system able to handle and respond to challenges in the digital age.

5 Deepfakes and the Criminal Process: Investigation to Trial

5.1 Deepfakes at the Investigation Stage

The implications for the criminal process are far more far-reaching than a mere matter of admissibility of a particular recording. Synthetic media have the capacity to impact a criminal case at its earliest stages, with subsequent challenges and difficulties arising during the prosecution and trial. Indeed, a fabricated recording might be the genesis of an investigation, while an authentic recording might be thrown into doubt by the mere possibility of manipulation. Thus, rather than a singular test applied when the evidence has reached the court-room, the criminal process should build in protections at each stage.

The initial difficulty arises when a suspected deepfake comes to the attention of an investigator. A video clip, photograph, or audio recording might be received by a law enforcement officer from social media, messaging apps, or another digital source, without any reliable indicator of its authenticity. If such a recording is accepted as authentic without investigation and analysis, it has the potential to send an investigation down the wrong track, especially when it attributes the recording to a particular individual (who is then identified as the alleged offender). The opposite danger of rejecting a potentially fabricated recording without a full review, even when it might be relevant to the offence, is equally serious.

Hence, the correct approach of investigators ought to be to preserve the recording and investigate its origins before concluding its authenticity. Properly documented searches and seizures of electronic devices should also continue. The BNSS provisions which require that certain searches and seizures be recorded by way of audio-video electronic recording can be considered an example of how technology may be used to strengthen procedural safeguards.[19] Digital evidence's vulnerability to modification during copying, transmission, or file conversion necessitates that investigators maintain a clear record of when and where the item was obtained, from whom or which device, how it was stored, and by whom and when it was later accessed or handled. This forms a chain through which the evidence can subsequently be examined by the courts.

5.2 Forensic Verification and Evidentiary Challenges

When the authentic status of digital media becomes disputed, examination by forensic experts begins to become crucial. Detecting a deepfake is likely to require inspection of the underlying file itself, the context in which the original was recorded, and the method of its creation and dissemination. Investigators may be expected to compare the recording under suspicion with an original, analyse metadata, and identify inconsistencies or tell-tale marks.[20]

Also, the very use of fabricated media may be of some worry for the prosecution. For instance, where recording of the statement is of importance in terms of the conviction, it would only be a natural step

[19] Bharatiya Nagarik Suraksha Sanhita, § 105.
[20] Bharatiya Sakshya Adhiniyam, § 39.

for defence lawyers to attack the authenticity of the evidence and perhaps the prosecution would be expected to prove that the recording in question in actual fact accurately documents what it claims to record. The prosecution should not take advantage of apparent and natural realism and will find that the judge has to strike a balance between the admissibility, authenticity and probative value of the recorded media. We should not, just because synthetic media exists, abandon core principles out of ease.[21]

5.3 Impact on Fair Trial and Judicial Assessment

A fabricated recording has the potential to create guilt in the public's imagination through sight before genuine evidence of fact exists and this makes fabricated audio and video the gravest threat of them all. Once disseminated, even after its spurious origin has been discovered, significant damage to the accused's reputation may linger. In the context of the criminal trial itself, reliance on inaccurate synthetic material may serve to compromise proceedings on the grounds of fairness; the constitutional right to a fair trial guarantees not only protections from potentially unfair judicial procedures, but also protects genuine evidence from being dismissed based on unsupported claims of artificial construction.[22]

Five key issues arise from this analysis.

i. Firstly, that visual or auditory fidelity cannot reliably be considered an indicator of authenticity.
ii. Secondly, that the existing rules governing electronic evidence did not anticipate the rise of fabricated media and hence do not provide a framework specifically geared to verifying content authenticity.
iii. Thirdly, that while forensic techniques for detection are now developing apace, such analysis cannot be regarded as conclusive.
iv. Fourthly, that fabricated evidence is not only a danger in the sense of being admitted and having undue weight assigned to it, but that genuine evidence, especially if of a technical nature, may also wrongly be excluded as if it were artificial.
v. Fifthly, and finally, that combating deepfakes cannot merely be the province of the courts. Indeed, the investigators preparing cases, the forensic scientists who will analyse them, prosecutors who will guide them and defence attorneys who will test them will have to acquire a far greater capacity to deal with synthetic media and the challenges it poses to veracity.

The criminal process, therefore, cannot simply take technology's word for it. Instead, the investigation and adjudication of cases involving digital material necessitates a layered approach – involving careful preservation of original material, rigorous forensic analysis where necessary, expert opinions that articulate limitations as clearly as conclusions, judicial assessment of content that considers provenance over mere appearance, the seeking of corroborative evidence, and an increased awareness from all participants of the threat posed by synthetic evidence. Such a method is better suited to allowing courts to maintain openness to the truth contained in evidence that has been recorded by technology, without necessarily accepting such a recording at face value merely on account of its convincing appearance.

---

[21] *Id.*

[22] Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473.

6 Enhancing India's Response: Legal, Forensic and Procedural Safeguards

6.1 Legal and Evidentiary Safeguards

The threat posed by deepfakes cannot be contained by simply adding new items to the cybercrimes list; the far more urgent need is to strengthen the means by which digital content is authenticated, inspected and presented before a court. What India needs is a holistic response consisting of corroborative protections, forensic capability, and institutional readiness. While the Bharatiya Sakshya Adhiniyam, 2023, already codifies evidence of electronically and digitally recorded records along with procedures for their admissibility (production and attestation),[23] deepfakes go beyond mere examination and require the court to ask if the recorded event and the content depicted actually occurred and is true in fact. Accordingly, specific principles of admissibility may be framed for deepfakes, especially when specific claims are made regarding artificial production of visual and audio evidence.

These principles aim not to indiscriminately bar deepfakes, but when authenticity is fairly put in issue and the recording is essential to the court's finding of facts, the court must have the power to demand some degree of forensic corroboration before admitting it as substantive evidence.

6.2 Forensic and Institutional Capacity

The need for specialized examination in deepfake cases cannot, therefore, be overstated. India requires adequate forensic labs and skilled personnel capable of analysing synthetic audio, still images, and video. Existing cyber-forensic facilities provide a base, but capacity should be expanded and made more accessible to various investigating agencies.[24]

Forensic reports should also, as much as possible, offer not only a simple "real" vs. "fake" verdict but also the underlying methodology and its limitations. Because synthetic content detection techniques evolve with generation techniques, judges must be able to assess the reliability of the methods employed in individual cases.

6.3 Preservation, Provenance and Stakeholder Training

Investigators should attempt to preserve original digital material whenever possible, in preference to relying solely on screen captures or forwarded versions. It is important to systematically document provenance, device of collection, relevant metadata, and subsequent handling of the digital content. The BSA provisions related to electronic evidence and hash values are helpful to confirm the integrity of preserved content but they should not imply that non-alteration after collection equates to original authenticity.[25]

All parties engaged in the criminal process – judges, police, prosecutors, and defence lawyers – need increased awareness of deepfake technology and digital forensics. While judges need not become technicians, a general understanding of how synthetic content is generated, and its detection, is valuable for judging expert testimony. Similarly, investigators and prosecutors should have a trained eye to identify suspicious digital evidence and not solely treat realistic-appearing audio or visual media as de facto proof. Defence lawyers can then use this training to raise proper challenges to the authenticity of evidence when there is real doubt.

---

[23] Bharatiya Sakshya Adhiniyam, §§ 61–63.

[24] Central Bureau of Investigation, National Cyber Forensic Laboratory, Government of India.

[25] Bharatiya Sakshya Adhiniyam, 2023, § 63 & sched.

A balanced response should safeguard the interests of victims while respecting the rights of the accused. Deepfakes can be used to promote sexual abuse, harassment, impersonation and fraud, but fake incriminatory evidence can similarly prejudice an innocent party. It is now also possible to dismiss genuine digital evidence on the ground of suspected falsity. The most balanced approach is therefore to focus on verification rather than assumption.

7 CONCLUSION

The impact of deepfakes on the relationship between technology and criminal justice can be profound. While electronic evidence is already an integral part of crime investigation and trial process, synthetic media has added a completely different dynamic to our dependency on the basic tenet that if an image or video looks genuine, it probably is an authentic representation of an actual incident. Therefore, the issue is not only that data has been manipulated, but that reality, as presented by the data, is fake. The present framework of evidence law in India, particularly under the Bharatiya Sakshya Adhiniyam, 2023, lays down some fundamentals regarding the admissibility and proving of electronic and digital evidence. However, this article aims to show that the standard requirements of the genuineness and integrity of digital evidence are not in and of themselves adequate enough to deal with the problem created by advanced synthetic material. A deepfake, while being sound, can be fake at the source. This creates an important difference between digital soundness and factual authenticity. The influence of the deepfakes affects several processes in the criminal trial, including the investigation, the scientific analysis of digital information, prosecution, the court's decision to judge the very authenticity and also the undue exclusion of genuine evidence out of fear that it might be synthesized. This is referred to as the liar's dividend, which will further reduce credibility. Hence, to address this problem there has to be an approach comprising reliable preservation measures, chains of custody, scientific expertise, expert opinion, judiciary, and corroborative proofs. Moreover there must be a focus on investing in infrastructure of cyber security, scientific experts, and well-trained investigators, prosecutors and judges. In essence deepfakes do not make electronic and digital evidence inadmissible; rather, they emphasize the need for proving. The criminal justice system in India now needs an approach by which it can clearly distinguish what appears real from what is factually authentic. The system must adapt, evolve, and embrace this. Thus the approach to digital evidence in India in the coming time should rely not on what is visibly "seen" and "heard" but on what can be effectively substantiated and proved in the case.

* * *

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.5

CHAPTER 5

Reading the Guilty Mind: Neurotechnology, Brain Data and the Evidentiary Determination of Mens Rea

DEBASRITA CHOUDHURY*

Mens rea is typically ascertained by the criminal law using external evidence, such as conduct, communications, motive, admissions, surrounding circumstances, expert testimony, and other factors from which a court may deduce the accused's mental state. This conventional evidential model is under threat from the rapid advancements in neurotechnology. It is possible that information from the brain itself may be used in criminal prosecutions thanks to techniques that can measure cerebral activity and identify patterns linked to recognition, decision-making, attention, and other cognitive functions. This poses an important question: can brain data prove legally significant mental states like knowing, recklessness, or intention?

With a focus on Indian criminal law, this study investigates the evidential and constitutional ramifications of employing neurotechnology to establish mens rea. It contends that while neurological evidence may be significant to criminal liability, a legally liable mental state cannot be equated with the presence of brain activity. The study highlights the inferential difference between legal proof of purpose or knowing and biological evidence of brain activity. In light of Selvi v. State of Karnataka (AIR 2010 SC 1974) and the Supreme Court's acknowledgment of privacy as a basic right in K.S. Puttaswamy v. Union of India (AIR 2017 SC 4161), it also explores the implications of brain-based evidence under Articles 20(3) and 21 of the Indian Constitution. Additionally, the research assesses the evidentiary framework of the Bharatiya Sakshya Adhiniyam, 2023.

According to the paper's cautious admissibility paradigm, neurotechnological evidence should be regarded as expert testimony that is subject to stringent scientific validation, individualized dependability, constitutional protections, and judicial review. Most significantly, mens rea shouldn't be independently established by brain data. Neurotechnological evidence that has been properly validated may, at most, support other evidence from which a court derives a legal conclusion on the accused's mental state.

Keywords: neurotechnology, mens rea, criminal liability, brain data, and evidential law

1 INTRODUCTION

Actus non facit reum nisi mens sit rea. It is the most basic tenet of criminal jurisprudence that the crime and the associated guilty mental state are the two fundamental prerequisites for criminal liability, both *actus* as well as *mens rea*. Criminal law typically distinguishes between the levels of associated intellect or mental state such as intention, knowledge, negligence or recklessness. And even though the exact definition of mens rea varies depending on the offence, the very obvious issue arises that the law is unable to directly see the mind, or detect the exact mental state during a crime.

Whether an accused deliberately ignored a risk, or had the requisite cognition of the foreseeable outcome, or meant to cause death, all these eventualities are determined by several indirect inputs to ascertain the mental state during the causation of crime. Courts employ evidence to rebuild the accused's mental

* Academician.

state. Such a conclusion may be corroborated by the weapon used, the attack technique, previous criminal activity, behaviour both before and after the offence, motive, and surrounding circumstances.

Our age-old method of evidence backed assertions of mental state faces a potentially revolutionary threat with the advent of neurotechnology. With the use of technologies like Electroencephalography (EEG)[1], Functional Magnetic Resonance Imaging (fMRI)[2], brain-computer interfaces, and other neural recording or decoding systems, modern neuroscience can quantify various aspects of brain activity and cognitive performance. Neural correlates of intention, awareness, risk perception, decision-making, and self-control have all been studied. As a result, researchers have started questioning whether neuroscience could give judges information directly related to criminal liability.

The point of contention is not whether research and innovation has caught up with the prolific intricacies of the human brain and what mapping such data could mean for the medical as well as the legal field. Rather, the question is, can it be relied upon solely to satisfy the legal nuances of attributing criminal liability? The basic difference between civil and criminal liability, as is taught even in imparting primary legal education, is that civil liability focuses on a preponderance of probability, that causation naturally may follow. However, a criminal liability is a much heavier burden and therefore requires the court to be satisfied that the accused is guilty beyond reasonable doubt. This standard extends to the preliminary requisite of mens rea, the act and therefore its consequences. At this juncture, can an innovative brain measurement prove a legally defined mental state with absolute certainty, to be held up in courts?

This distinction is crucial. A neural pattern linked to decision-making should not immediately prove that the accused legally intended a specific outcome. Similarly, awareness of an act's unlawful nature may not be essentially established by a brain pattern linked to recognition. There are multiple levels of interpretation involved in the process, from neurological observation to legal decision.

This study makes the case that constitutional and evidentiary restrictions should apply to the use of brain data in India if it is used as proof of mens rea.

2 CORRELATING MENS REA AS A LEGAL CONCEPT AND BRAIN DATA ANALYSIS

Mens rea is a normative legal concept. Under criminal law, certain behaviours are viewed as more culpable than others, hence it gives importance to mental states. It is possible for two persons to carry out nearly identical physical acts while having distinct mental states, which could lead to different criminal outcomes. The mental connection to the act is essentially different, but the physical outcome is comparable. As a result, law poses queries such as, did a defendant plan the outcome? Was the accused aware of the likelihood of the outcome? Was the behaviour unintentional or did the accused have the legal capacity to comprehend the nature or ramifications of the behaviour? The importance of mental states is not eliminated by the Bharatiya Nyaya Sanhita, 2023. Concepts like intention, knowledge, criminal intention, and certain types of awareness are nevertheless used in its general exclusions and offense-specific restrictions. For instance, Section 22[3] addresses the actions of a person of unsound mind who is unable to understand

[1] J.W. BRITTON, L.C. FREY ET AL., ELECTROENCEPHALOGRAPHY (EEG): AN INTRODUCTORY TEXT AND ATLAS OF NORMAL AND ABNORMAL FINDINGS IN ADULTS, CHILDREN, AND INFANTS Introduction (2016), https://www.ncbi.nlm.nih.gov/books/NBK390346/ (last visited Aug. 26, 2026).

[2] G.H. Glover, *Overview of Functional Magnetic Resonance Imaging*, 22 NEUROSURG. CLIN. N. AM. 133–39, vii (Apr. 2011).

[3] Bharatiya Nyaya Sanhita, 2023, § 22.

the nature of the conduct or that it is illegal, whereas Section 18[4] recognizes accidents when an act is undertaken without criminal intention or awareness under specific circumstances. In specific circumstances, Indian criminal adjudication still requires courts to ascertain mental state. Therefore, neurological data may go a long way in predicting, with some accuracy, the legal query thus raised.

The field of neuroscience has created techniques for examining the anatomy and functioning of the brain. EEG captures scalp electrical activity. fMRI is frequently employed as an indirect measure of brain activity and measures changes related to blood oxygenation.[5] Neural signal recording and interpretation for technical and medicinal applications is becoming more and more possible. Brain activity has been studied in relation to intention, voluntary action, risk perception, consciousness, and self-control. Research in existence acknowledges that this kind of data could someday have an impact on legal issues related to criminal liability.[6] The question of whether distinct psychological and neurological processes correspond to various legally relevant mental states is one significant area of research.

For instance, established literature deliberates on the matter whether legal distinctions like "knowing" and "reckless" correspond to sufficiently different mental and neural states to make brain-based detection possible.[7] However, admissibility and probability of being admissible are not synonymous. Mere suggestion of 'knowing' the possibility of a specific outcome does not negate decades of criminal jurisprudence in establishing a mind state.

It raises a question whether neural correlation is verifiable by law. The temptation to accept scientific measurement as the same as certainty is the biggest risk in the legal system.

Hypothetically, if a pattern statistically linked to conscious decision-making is found by a brain-scanning device, the prosecution may contend that because the neural pattern was present, the accused deliberately desired the outcome. It is not always the case that this conclusion follows. The inferential chain is far longer in reality.

Neural activity presupposes cognitive function which in turn leads to a psychological state, the cumulative sum of which is much greater than the legally significant mental state that leads to criminal responsibility.

A brain scan may reveal activity related to making decisions which doesn't necessarily prove what the individual chose, why they decided it, or whether the decision meets the legal definition of intention. The types of individual mental states that criminal law requires judges to ascertain are still difficult to infer from current research.[8]

A significant challenge is that a large portion of neuroscientific research is conducted at the group level. A study may find a statistical correlation between a psychological task and a specific brain pattern in a group of volunteers. However, a criminal court does not decide whether a group was mentally ill. It involves ascertaining if the accused had the mental condition necessary for this specific offence at the moment. Thus, it is important to distinguish between group-level correlation and individual-level inference. When it comes to determining whether a specific defendant had a certain mental state at the time

---

[4] *Id.* § 18.

[5] G.H. Glover, *Overview of Functional Magnetic Resonance Imaging*, 22 NEUROSURG. CLIN. N. AM. 133–39, vii (Apr. 2011).

[6] David L. Faigman, John Monahan & Christopher Slobogin, *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 U. CHI. L. REV. art. 1 (2014), https://chicagounbound.uchicago.edu/uclrev/vol81/iss2/1 (last visited Sept. 5, 2026).

[7] Owen D. Jones, Read Montague & Gideon Yaffe, *Detecting Mens Rea in the Brain*, 169 U. PA. L. REV. 1 (2020), https://scholarship.law.vanderbilt.edu/faculty-publications/1293 (last visited Sept. 3, 2026).

[8] S.J. Morse, *Neurolaw: Challenges and Limits*, 197 HANDBOOK OF CLINICAL NEUROLOGY 235–50 (2023), https://doi.org/10.1016/B978-0-12-821375-9.00003-7 (last visited Sept. 5, 2026).

of a certain criminal act, an algorithm might be very effective at finding trends in a research population, but it might not be sufficiently dependable.[9]

3 THE EVIDENTIARY STATUS OF NEUROLOGICAL DATA UNDER INDIAN LEGAL FRAMEWORK

The current Indian framework for evidence admissibility and proof is provided by the Bharatiya Sakshya Adhiniyam, 2023. Sections 61-63[10] deal with electronic or digital documents and their evidentiary value, whereas Section 39[11] deals with expert opinion.

Evidence from neurotechnology does not fall into a conventional category. One possible digital record is a brain scan. Expert opinion may be included in the scan's interpretation. Algorithmic inference may also be used in a neural-decoding system. Consequently, a court ought to make at least three distinctions: unprocessed neural data, the data's scientific interpretation and the intended legal inference derived from the interpretation. A technological record might be the first. The second could be covered by expert testimony. The court still has to decide on the third. The scientific meaning of a brain signal can be explained by an expert. However, the final legal question of whether the accused had mens rea should not be decided by the expert. Neurological data may be pertinent, but this does not imply that it is trustworthy enough to support a criminal conviction. Therefore, a robust evidentiary framework should differentiate between individual dependability, scientific validity, procedural integrity, relevance, and alignment with constitutional values.

Before relying on any neurological technology as evidence, the prosecution ought to establish primarily the rate of errors of the technology, the general approach employed, whether the outcome is repeatable, if the technology was correctly adjusted given outliers, whether the outcome can be skewed by environmental influences, whether independent testing has been done on the algorithm. Most importantly, it ought to be established convincingly whether the outcome is applicable to the specific accused. As neuroscience data can seem more authoritative than its true probative value justifies, this strategy is especially crucial. Exaggerated assumptions about the legal significance of neuroscience have been cautioned against by neurolaw scholars on several occasions.[12]

3.1 Can Brain Data Be Self-Incriminating under Article 20(3)?

According to Article 20(3)[13] no one accused of a crime may be forced to testify against himself.

Self-incrimination has always been understood in terms of forced statements. The leading judgement on the facets of self-incrimination, State of Bombay v. Kathi Kalu Oghad,[14] has stipulated that providing physical evidence such as handwriting and blood samples will not come under the ambit of this Article, thus exempting certain categories of evidence from this defence. The recent judgement of Ritesh Sinha

---

[9] Christopher Slobogin, *Neuroscience Nuance: Dissecting the Relevance of Neuroscience in Adjudicating Criminal Culpability*, 4 J.L. & BIOSCIENCES 577–93 (Dec. 2017), https://doi.org/10.1093/jlb/lsx033 (last visited Sept. 5, 2026); David L. Faigman, John Monahan & Christopher Slobogin, *Group to Individual (G2i) Inference in Scientific Expert Testimony*, 81 U. CHI. L. REV. art. 1 (2014), https://chicagounbound.uchicago.edu/uclrev/vol81/iss2/1 (last visited Sept. 5, 2026).

[10] Bharatiya Sakshya Adhiniyam, 2023, §§ 61, 62, 63.

[11] *Id.* § 39.

[12] E. Aharoni, C. Funk, W. Sinnott-Armstrong & M. Gazzaniga, *Can Neurological Evidence Help Courts Assess Criminal Responsibility? Lessons from Law and Neuroscience*, 1124 ANNALS N.Y. ACAD. SCI. 145–60 (2008), https://doi.org/10.1196/annals.1440.007 (last visited Sept. 5, 2026).

[13] INDIA CONST. art. 20, cl. 3.

[14] State of Bombay v. Kathi Kalu Oghad, 1961 AIR 1808.

v. State of Uttar Pradesh[15] confirmed that voice samples fall within this same category. However with advanced neurotechnology, the difference between "speaking" and "revealing" is called into question. Should information obtained by a procedure from a person's brain be regarded as testimonial or tangible evidence if the individual does not need to intentionally disclose it, or more so, consent to it?

In Selvi v. State of Karnataka[16] the Supreme Court addressed an earlier iteration of the same issue.

Narco-analysis, polygraph tests, and Brain Electrical Activation Profile (BEAP), often known as brain-mapping techniques, were all invoked in this case. According to Article 20(3)[17] and Article 21[18] the Supreme Court considered the involuntary use of these methods to be unconstitutional. As the Court acknowledged that the protection against self-incrimination cannot be limited to the actual act of speaking, its ruling is especially noteworthy.

Future neurotechnology will be greatly impacted by the constitutional principle that emerged from Selvi. In the past, Indian constitutional law has made a distinction between the gathering of tangible evidence and forced testimony. Typically, a forced statement provides more insight into a person's thoughts than a fingerprint, blood sample, or handwriting specimen. This difference is complicated by brain data. The State may contend that a brain device is comparable to physical evidence if it only records physiological data.[19]

However, if the technology decodes acknowledgment of a specific incident, the person's memories, recognition of a fact, the goal, a sentimental reaction, choices, or other data stored in the individual's mental processes, the claim that the evidence is only physical is significantly undermined. Possibility of self-incrimination increases with the degree to which a technology directly incorporates personal knowledge. *Selvi* is more than just a case using outdated investigative methods. Its treatment of mental privacy and forced extraction of personal knowledge has deeper constitutional relevance. As a result, the Supreme Court's ruling can offer a constitutional basis for examining technologies that are far more advanced than BEAP. In fact, it may be argued that a future technology that could decipher an individual's intention would pose a more significant issue than a method that could just detect knowledge of particular information. Therefore, the constitutional concern should focus on the type of information being retrieved rather than just how it was obtained.[20]

3.2 The Qualms of Privacy

The Supreme Court acknowledged privacy as a fundamental right safeguarded by Article 21[21] and linked to the rights provided by Part III of the Constitution in the case of K.S. Puttaswamy v. Union of India.[22] Secrecy about communications or physical locations is not the only aspect of privacy. It also includes control over private facets of one's life, autonomy, and dignity.

The brain is intimately linked to contemplation, memory, purpose, sentiment, decision-making, identity, and awareness, thus making neural data a particularly sensitive category. As a result, brain data can

[15] Ritesh Sinha v. State of Uttar Pradesh, (2019) 8 SCC 1.
[16] Selvi v. State of Karnataka, (2010) 7 SCC 263.
[17] INDIA CONST. art. 20, cl. 3.
[18] *Id.* art. 21.
[19] Javier Escobar Veas, *Brain-Reading Technologies and the Right Against Self-Incrimination: A Challenge for the Distinction Between Testimonial and Real Evidence*, 26 GERMAN L.J. 683–701 (2025).
[20] *Id.*
[21] INDIA CONST. art. 21.
[22] Justice K.S. Puttaswamy (Retd) v. Union of India, (2017) 10 SCC 1.

be entitled to a higher level of security than regular personal data. This issue is the focus of the new notion of mental privacy, often known as cognitive liberty. DNA may disclose biological traits. However, what a person remembers, recognizes, desires, fears, or decides may be revealed by neural data. The latter may be far more invasive of the person's mental autonomy.[23]

Traditionally, the State uses behaviour observation to look into crimes. By penetrating the evidentiary realm of mind itself, neurotechnology may enable research but it also raises a constitutional question that could be referred to as the "last private frontier" of personal autonomy: should people be allowed to maintain a protected space where their thoughts and cognitive processes cannot be extracted by the state unless there are extremely compelling circumstances?

3.3 Facing Overbearing Criminal Jurisprudence

Neurotechnology has a base issue with criminal law, even if constitutional problems are resolved. It is up to the prosecution to prove guilt beyond a reasonable doubt. Hypothetically, if a neural technology can identify a specific cognitive state with 90% accuracy it might look impressive. However, a 90% accuracy rate does not prove beyond a reasonable doubt that the accused person had the pertinent mental state. Typically, the criminal justice system doesn't inquire: "Is this technology generally accurate?" It poses: "Has the prosecution met the necessary standard to prove this accused's guilt?" This leads to a significant epistemological issue. Probabilistic evidence could be produced by a technology but a legal conclusion is necessary in criminal law. Therefore, the court must refrain from turning statistical certainty into criminal certainty.[24]

The most appealing and challenging argument for neurotechnology is intention. Neuroscience can research the cognitive mechanisms involved in intentional action. Brain activity linked to motor skills, consciousness, agency, and decision-making has already been studied.[25]

However, the legal definition of intention is situational. It takes into account intention, causation, knowledge of actions and foresight of their consequences, acts of self-defence and a lot more factors. Information about cognitive activity could be obtained from a brain signal. However, the court must take the surrounding facts into account in order to classify that conduct legally. A neurological scan might not be as probative as the weapon used, distance between accused and victim, words spoken, number of injuries, prior threats, and subsequent behaviour. Thus, contextual reasoning should not be replaced by neurotechnology. It might, at most, complement it.[26]

Neuro-determinism poses another conceptual risk. The normative underpinnings of criminal liability may be threatened by such a view. People are not punished by the law just because they have a specific brain anatomy. Additionally, it shouldn't penalize people because an algorithm finds a brain tendency linked to antisocial behaviour. There is a basic difference between tendency and responsibility. Without committing any crimes, a person may have a neurobiological trait that is statistically linked to impulse or pre-disposition to violence.[27]

---

[23] Francis X. Shen, *Neuroscience, Mental Privacy, and the Law*, 36 HARV. J.L. & PUB. POL'Y 653 (2013), https://scholarship.law.umn.edu/faculty_articles/599 (last visited Sept. 7, 2026).

[24] N. Eastman & C. Campbell, *Neuroscience and Legal Determination of Criminal Responsibility*, 7 NATURE REVS. NEUROSCIENCE 311–18 (2006), https://doi.org/10.1038/nrn1887 (last visited Sept. 7, 2026).

[25] *Id.*

[26] Slobogin, *supra* note 9.

[27] Morse, *supra* note 8.

In a similar vein, a neurological condition might be important to an insanity defence or another legally recognized exception, but it wouldn't immediately absolve someone of blame. Neuroscience evidence may be pertinent to issues like involuntariness, mens rea, insanity, and mitigation, according to existing research, but its legal value varies greatly depending on what the evidence really shows.[28]

In some situations, the application of neuroscience may be more justifiable than the direct proof of mens rea. Acts of a person of unsound mind who, at the relevant time, is unable to understand the nature of the conduct or that what they are doing is improper or illegal are recognized as an exception under Section 22[29] of the BNS. In this case, neurological evidence could help the judge comprehend impairment of cognition and circumstances that impact logical comprehension. But even in this case, it's crucial to distinguish between medical diagnosis and legal culpability. Whether the statutory requirements are met is still a pertinent legal concern. Neurotechnology should therefore supplement legal judgment rather than replace it.[30]

3.4 Should Neurological Data Be Treated as a New Category of Evidence?

As neurotechnology advances, the current categories of evidence might become more and more insufficient. A few types of neuro-evidence could be distinguished using a conceptual classification. For example, *structural evidence* relates to the physical features of the brain, such as signs of neurological damage or anomalies. It may be primarily relevant to ability, illness, impairment. *Functional evidence* relates to brain activity patterns connected to specific cognitive functions. Since correlation does not always prove a specific mental state legally, its applicability to mens rea is less clear. *Decrypted cognitive evidence* is the most sensitive classification. Technology that can convert brain activity into data on memories, intentions, preferences, recognition, or other cognitive content would be included. The case for considering neural information as a type of testimonial material protected by the Constitution gets stronger as technology advances in its ability to decode cognitive content.[31]

4 The Way Forward for the Indian Legal System

With the advent of research in neurotechnology, admissibility of evidence as explained in 'neurolaw' may not be as far in the future as one would believe. Courts would have to deliberate on a progressive admissibility framework. Evidence proving neurological impairment should be handled differently from evidence proving the accused's intent to kill. The court ought to differentiate between brain structure data, physiological reactions, cognitive information decoded among others.[32]

The prosecution needs to prove a scientific foundation of the technology relied upon, known rates of error, repeatability, autonomous verification, dependability in practical situations, suitable control groups, validity at the individual level etc.[33]

---

[28] Vitacco et al., *Neuroscience and the Insanity Defense: Trying to Put a Round Peg in a Square Hole*, 5 Forensic Sci. Int'l: Mind & L. art. 100131 (2024).

[29] Bharatiya Nyaya Sanhita, 2023, § 22.

[30] Shen, *supra* note 23.

[31] U. Maoz & G. Yaffe, *What Does Recent Neuroscience Tell Us About Criminal Responsibility?*, 3 J.L. & Biosciences 120–39 (2015), https://doi.org/10.1093/jlb/lsv051 (last visited Sept. 7, 2026).

[32] Eastman, *supra* note 24.

[33] *Id.*

The court should take Articles 20(3)[34] and 21[35] into independent consideration when the acquisition of such data was not voluntary. When the technology gathers personal information instead of only documenting physical attributes, further caution should be used. Even admissible neurotechnology shouldn't be given decisive weight by default. The court ought to take into account firsthand evidence, documentary proof, electronic correspondence, forensic proof, circumstantial proof, expert testimony etc.[36]

The court must continue to make the final legal conclusion regarding mens rea. The *Corroboration, not Conviction Principle* is the idea put forth in this work. Until future scientific advancements show an exceptionally high degree of individualized reliability and the constitutional framework is correspondingly developed, neurotechnological evidence may support other evidence relevant to mens rea but should not independently establish criminal intention, knowledge, or recklessness. This acknowledges neuroscience's strengths and weaknesses. The intention should continue to be deduced from all of the evidence that is admissible under the law.

5 COMPARATIVE PERSPECTIVES AND THE EMERGENCE OF NEURO-RIGHTS

The growing controversy surrounding 'neuro rights' is especially pertinent. Technology access to the brain may necessitate legal safeguards beyond standard data protection.

Indian constitutional law acknowledges the unique relevance of involuntary information extraction from the mind, as demonstrated by the constitutional reasoning in *Selvi.*[37] *The constitutional* basis for safeguarding mental autonomy is further reinforced by *Puttaswamy's*[38] acknowledgment of privacy as a basic right. Criminal courts already draw conclusions about mental states from incomplete evidence, which is the best justification for neurotechnology. Excluding neuroscience completely could deprive courts of potentially helpful evidence if it can offer more information. Expert testimony that is probabilistic rather than definitive is frequently used by courts. Despite their limits, psychiatric evaluations, forensic science, and medical evidence can help courts.

Neurotechnology in the United States is at a crossroads between disjointed privacy regulations and strict evidentiary gates. Because thought-decoding has a significant mistake rate and is not widely accepted in science, courts frequently prohibit advanced neuroimaging, such as fMRI lie detection, under the Daubert or Frye criteria.[39] As a result of the lack of a federal framework, individual states like Colorado and California have taken the lead by specifically updating their consumer privacy laws to protect sensitive neural data from commercial exploitation. Meanwhile, the legal battleground is rapidly shifting toward mental privacy.[40]

In the UK, neurotechnology is viewed more as an interpretive medical tool than as an impartial "truth machine." UK courts easily accept structural brain scans to demonstrate organic brain injury under Civil

[34] INDIA CONST. art. 20, cl. 3.
[35] *Id.* art. 21.
[36] Neal Feigenson, *Brain Imaging and Courtroom Evidence: On the Admissibility and Persuasiveness of fMRI*, 2 INT'L J.L. CONTEXT 233, 234 (2006).
[37] Selvi v. State of Karnataka, (2010) 7 SCC 263.
[38] Justice K.S. Puttaswamy (Retd) v. Union of India, (2017) 10 SCC 1.
[39] Frye v. United States, 293 F. 1013 (D.C. Cir. 1923); Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993); https://www.expertinstitute.com/resources/insights/daubert-vs-frye-navigating-the-standards-of-admissibility-for-expert-testimony/ (last visited Sept. 8, 2026).
[40] Kristina Iliopoulos & Nancy L. Perkins, *Neural Data Privacy Regulation: What Laws Exist and What Is Anticipated?*, ARNOLD & PORTER (2025).

Procedure Rules and criminal standards, but they mostly rely on expert psychiatric testimony to convert those physical lesions into legal arguments about mental capacity or impulse control. British judicial scrutiny focuses on the impartiality and dependability of expert interpretation rather than on sophisticated lie detection, guaranteeing that neuro-psychiatric evidence stays simply necessary and pertinent to the court's comprehension.[41]

6 Conclusion

Before neurotechnological evidence is regularly made available to investigating agencies, India should think about creating a unique legal framework for it, including defining brain data separately due to its potential to expose cognitive content. It should outlaw forced cognitive extraction. Judicial monitoring ought to be required whenever neurotechnology is suggested in a criminal inquiry. Instead of depending solely on prosecution-retained specialists, courts should have access to independent scientific expertise. For a meaningful challenge, the accused should have access to the underlying data, methodology, algorithmic settings, and independent expert evaluation. India should prevent the right to a fair trial from being compromised by proprietary secrecy. Constitutional examination should become far more stringent as technology advances from tracking brain activity to decoding mental information.

Neuroscience itself is not the threat. Neuro-reductionism, the belief that a lawful mental state may be fully reduced to a quantifiable biological state, poses a threat. The moral underpinnings of criminal responsibility as well as evidentiary reasoning may be distorted by such reductionism.[42]

Knowledge is not inherently derived from a brain pattern. Mens rea is not always a cognitive process. The primary challenge is the inferential divide between criminal law and neuroscience. Neural activity and psychological processes may be correlated by neuroscience; however, criminal liability is contingent upon legally defined mental states in the context of specific behavior. Although the Bharatiya Sakshya Adhiniyam offers a framework for the possible consideration of scientific and digital evidence, issues of dependability, individualization, and constitutional validity are inextricably linked to technical admissibility. Therefore, the proper Indian strategy should neither deny neuroscience nor give it control over criminal decisions.

Instead of taking the place of the court's contextual evaluation of purpose and knowledge, brain data should typically be used as corroborating evidence. The tempting notion that technology can merely "read the guilty mind" should ultimately be rejected by criminal law. Determining what transpired inside a brain is not the only goal of criminal adjudication. It is to decide whether someone should be held criminally liable for an act based on legally established standards of culpability and proof. Any future Indian framework governing neurotechnological evidence must continue to centre around this distinction.

* * *

[41] P. Catley & L. Claydon, *The Use of Neuroscientific Evidence in the Courtroom by Those Accused of Criminal Offenses in England and Wales*, 2 J.L. & Biosciences 510–49 (2015), https://doi.org/10.1093/jlb/lsv025.

[42] M. Fallin, O. Whooley & K.K. Barker, *Criminalizing the Brain: Neurocriminology and the Production of Strategic Ignorance*, 14 BioSocieties 438–62 (2018), 10.1057/s41292-018-0135-y.

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.6

CHAPTER 6

Artificial Intelligence Enhanced Forensics and Child Rights Protection: Policy, Law and Ethics

YAMINI PATEL* AND DR. GRISHMA BHAVSAR**

The integration of Artificial Intelligence (AI) into forensic science is transforming criminal investigations across jurisdictions. In India, where child sexual abuse cases are prosecuted under the Protection of Children from Sexual Offences Act, 2012 (POCSO) framework, forensic evidence plays a decisive role in ensuring fair trials and safeguarding child rights. However, the adoption of Artificial Intelligence (AI) assisted forensic tools such as automated DNA analysis, pattern recognition, facial analytics, and digital evidence processing raises complex legal, ethical, and procedural concerns that remain insufficiently examined within the existing statutory structure.

It argues that while Artificial Intelligence (AI) has the potential to enhance accuracy, speed, and objectivity in evidence collection and analysis, the absence of clear legal standards, accountability mechanisms, and child sensitive protocols may risk procedural injustice and secondary victimization. Artificial Intelligence (AI) in forensic investigation can advance social justice only when governed by a rights-based, child centric legal framework.

Keywords: Artificial Intelligence, Forensic Evidence, Child Rights, Algorithmic Accountability, Social Justice.

1 INTRODUCTION

Artificial Intelligence (AI) has emerged as one of the most significant technological developments in contemporary forensic science, transforming the manner in which criminal investigations are conducted and forensic evidence is examined. Advances in machine learning, computer vision, and predictive analytics have enabled forensic laboratories to analyse complex biological, digital, and multimedia evidence with greater speed and precision than conventional methods.[1] Artificial Intelligence (AI) assisted tools are increasingly employed in DNA mixture interpretation, facial comparison, fingerprint analysis, image enhancement, voice recognition, and cyber forensic investigations, thereby improving the efficiency of evidence identification and reducing investigative delays.[2] These developments have attracted considerable attention from governments, forensic institutions, and law enforcement agencies across the world because of their potential to strengthen criminal justice administration while addressing the growing volume of digital evidence.[3] Nevertheless, the integration of Artificial Intelligence (AI) into forensic practice is not merely a technological advancement; it also raises fundamental questions regarding transparency,

* PhD Research Scholar at GLS University, Ahmedabad, Gujarat, India.
** Assistant Professor at GLS University, Ahmedabad, Gujarat, India.

[1] NATIONAL INSTITUTE OF JUSTICE, ARTIFICIAL INTELLIGENCE AND FORENSIC SCIENCE: A FRONTIER FOR INNOVATION IN CRIMINAL JUSTICE (U.S. Dep't of Justice 2024).

[2] INTERPOL, TOOLKIT FOR RESPONSIBLE AI INNOVATION IN LAW ENFORCEMENT (2023).

[3] EUROPOL, FACING REALITY? LAW ENFORCEMENT AND THE CHALLENGE OF DEEPFAKES (2022).

explainability, accountability, algorithmic bias, and the admissibility of Artificial Intelligence (AI) generated evidence before courts.[4] The increasing reliance on automated decision support systems therefore requires a careful examination of whether technological innovation remains compatible with constitutional guarantees, procedural fairness, and the rule of law.[5]

These concerns assume greater significance in cases involving child sexual abuse, where the search for scientific accuracy must be balanced with the protection of children's rights and the preservation of their dignity throughout the criminal justice process. The Protection of Children from Sexual Offences Act, 2012 (POCSO Act) establishes a specialised legal framework that recognises children as vulnerable participants in criminal proceedings and mandates child friendly procedures during investigation, evidence collection, and trial.[6] Recent legislative developments, including the Bharatiya Sakshya Adhiniyam, 2023 and the Digital Personal Data Protection Act, 2023, further underscore the importance of authenticity, privacy, and responsible data governance in an increasingly digital justice system.[7] Despite the growing use of Artificial Intelligence (AI) in forensic investigations, limited legal scholarship examines its impact on child protection, evidentiary fairness, privacy, and ethical standards. Against this backdrop, this study examines the legal, policy, and ethical implications of using Artificial Intelligence (AI) in forensic investigations involving child sexual abuse. It argues that while Artificial Intelligence (AI) possesses considerable potential to improve forensic accuracy and strengthen child protection, its legitimacy ultimately depends upon transparent governance, human oversight, judicial accountability, and a regulatory framework that places the best interests of the child at the centre of technological innovation.[8]

2 LEGAL FRAMEWORK GOVERNING CSA INVESTIGATIONS IN INDIA

The investigation of child sexual abuse (CSA) in India is governed by an interrelated statutory framework that seeks to balance effective criminal investigation with the protection of children's rights. At the centre of this framework is the Protection of Children from Sexual Offences Act, 2012 (POCSO Act), which establishes a child centric system of investigation and trial. Unlike ordinary criminal legislation, the Act not only defines offences against children but also prescribes safeguards to minimise secondary victimisation.

It mandates child friendly procedures for recording statements, conducting medical examinations, preserving confidentiality, and ensuring expeditious trials before Special Courts.[9] Consequently, forensic investigation under the POCSO Act is not merely an evidentiary exercise; it is an integral component of child protection that must uphold the dignity, privacy, and best interests of the child throughout the investigative process.[10]

The procedural framework governing criminal investigations is provided by the Bharatiya Nagarik Suraksha Sanhita, 2023 (BNSS), which authorises investigating agencies to collect scientific evidence,

---

[4] FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016).

[5] INDIA CONST. arts. 14, 21; Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1.

[6] Protection of Children from Sexual Offences Act, 2012.

[7] Bharatiya Sakshya Adhiniyam, 2023; Digital Personal Data Protection Act, 2023.

[8] UNESCO, RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE (2021); OECD, OECD PRINCIPLES ON ARTIFICIAL INTELLIGENCE (2019).

[9] Protection of Children from Sexual Offences Act, 2012.

[10] Convention on the Rights of the Child arts. 3, 16, 19, Nov. 20, 1989, 1577 U.N.T.S. 3; INDIA CONST. arts. 14, 21.

conduct medical examinations, seize relevant material, and forward evidence for forensic analysis.[11] While these provisions recognise the increasing importance of scientific methods in criminal investigations, they do not expressly regulate the use of Artificial Intelligence (AI) in forensic decision making. As a result, investigators may employ AI assisted technologies without statutory standards governing their validation, reliability, or accountability, creating uncertainty regarding the legal status of algorithm assisted forensic findings.

The evidentiary framework is contained in the Bharatiya Sakshya Adhiniyam, 2023 (BSA), which recognises scientific opinion, expert testimony, and electronic records as admissible evidence, subject to the requirements of authenticity and reliability.[12] However, the growing use of AI assisted forensic tools introduces new evidentiary challenges. Unlike conventional expert analysis, AI systems often generate conclusions through complex computational models that are not readily explainable before a court. This creates tension between technological efficiency and the principles of transparency, judicial scrutiny, and the accused's right to effectively challenge expert evidence during trial.

The increasing prevalence of cyber enabled child sexual offences further expands the role of digital evidence in criminal investigations. Online grooming, child sexual abuse material (CSAM), encrypted communications, and social media interactions frequently require specialised cyber forensic examination. The Information Technology Act, 2000 provides the legal basis for recognising electronic records and investigating cyber offences, thereby facilitating the collection and preservation of digital evidence.[13] Although AI applications such as automated image classification, metadata reconstruction, and behavioural pattern analysis are increasingly used in digital forensics, their deployment remains largely governed by general evidentiary principles rather than a dedicated statutory framework.

The protection of children's personal information has acquired greater significance following the enactment of the Digital Personal Data Protection Act, 2023 (DPDP Act). Investigations involving child victims routinely require the processing of highly sensitive personal data, including medical records, biometric information, forensic images, and electronic communications. The DPDP Act incorporates principles such as lawful processing, purpose limitation, data minimisation, and enhanced protection for children's personal data.[14] These safeguards assume particular importance where AI systems retain, analyse, or continuously learn from sensitive datasets, raising concerns regarding privacy, data security, and secondary use of information beyond the original investigative purpose.

Although India's legal framework recognises the growing importance of scientific and digital evidence, none of the existing statutes specifically regulates AI generated forensic outputs. There are no statutory standards governing the validation of forensic algorithms, independent auditing of AI systems, disclosure of algorithmic methodologies, or judicial assessment of AI assisted expert evidence. Similarly, neither the POCSO Act nor the Bharatiya Sakshya Adhiniyam prescribes procedural safeguards for evaluating the reliability, explainability, or accountability of AI generated forensic reports. This regulatory silence creates a significant gap between technological innovation and legal preparedness.

Unless supported by clear legal standards and robust institutional oversight, the increasing reliance on AI assisted forensic technologies may undermine evidentiary fairness, procedural transparency, and

[11] Bharatiya Nagarik Suraksha Sanhita, 2023.
[12] Bharatiya Sakshya Adhiniyam, 2023.
[13] Information Technology Act, 2000.
[14] Digital Personal Data Protection Act, 2023.

the child centred objectives that underpin India's legal framework for the protection of children from sexual offences.[15]

3 CASE LINKED ILLUSTRATIONS FROM INDIA

Indian courts have increasingly recognised the importance of forensic and digital evidence in the investigation and prosecution of sexual offences, particularly those involving children. Although no reported decision directly examines the admissibility of Artificial Intelligence (AI) assisted forensic evidence in POCSO cases, several judicial pronouncements demonstrate the growing reliance on scientific methods and digital technologies in criminal investigations. These decisions provide the legal foundation for evaluating the future role of AI in forensic investigations while also highlighting the absence of specific judicial standards governing AI generated evidence.

a) *Mukesh v. State (NCT of Delhi) (2017)*

Although not a POCSO case, the Supreme Court relied extensively on DNA forensics and scientific evidence, demonstrating how decisive forensic credibility can be in sexual offence trials.[16]

b) *State of Himachal Pradesh v. Rajesh Kumar (2019)*

The Court emphasized the evidentiary value of DNA profiling in a POCSO matter, showing how scientific evidence can outweigh testimonial inconsistencies.[17]

c) *Alakh Alok Srivastava v. Union of India (2018)*

Directions were issued regarding online child sexual abuse material and the need for technological mechanisms to detect and prevent circulation, highlighting the growing intersection of technology and child protection.[18]

d) Digital Evidence in Online Grooming and CSA Material Cases

Courts have had to assess chat records, metadata, and device forensics in POCSO prosecutions, indicating a shift toward digital/AI assisted evidentiary reliance.[19]

These cases show courts trusting forensic and digital outputs, even though AI governance norms are absent.

4 AI (ARTIFICIAL INTELLIGENCE) APPLICATIONS IN CONTEMPORARY FORENSICS

Artificial Intelligence is reshaping forensic science by enabling faster processing, pattern recognition, and analytical interpretation of complex evidence. In CSA investigations, where time sensitive biological traces and large volumes of digital material are common, AI assisted tools increasingly support investigators and forensic experts in managing evidentiary complexity.[20] One of the most important applications is in DNA analysis and probabilistic genotyping. Traditional DNA interpretation becomes difficult when

[15] UNESCO, RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE (2021); NIST, AI RISK MANAGEMENT FRAMEWORK (AI RMF 1.0) (2023).

[16] Mukesh v. State (NCT of Delhi), (2017) 6 SCC 1.

[17] State of Himachal Pradesh v. Rajesh Kumar @ Raju, (2019) 14 SCC 268.

[18] Alakh Alok Srivastava v. Union of India, (2018) 17 SCC 291.

[19] *See, e.g.*, Shinoj v. State of Kerala, Crl. A. No. 2389 of 2024, 2025 LiveLaw (Ker) 654 (Ker. HC Oct. 17, 2025) (chat records in a POCSO appeal; redacted conversations inadmissible even with a Section 65B certificate); Just Rights for Children Alliance v. S. Harish, 2024 INSC 716 (computer forensic analysis report on the accused's mobile phone).

[20] Christopher Rigano, *Using Artificial Intelligence to Address Criminal Justice Needs*, NIJ J. No. 280 (Nat'l Inst. of Justice, U.S. Dep't of Justice 2019), NCJ 252038.

samples are degraded, contaminated, or mixed. AI driven software assists experts in interpreting complex DNA mixtures and generating statistical probability assessments with improved speed and consistency. In CSA cases, where delayed reporting often weakens biological evidence, such tools can significantly influence evidentiary outcomes.[21]

AI is also used in image and video enhancement. In cases involving child sexual abuse material (CSAM), CCTV footage, or low-quality visuals, AI algorithms can enhance clarity, reconstruct blurred frames, and detect relevant visual patterns. This allows recovery of usable evidence from otherwise compromised material.[22] Another emerging use is facial comparison and pattern recognition. AI systems compare facial features across datasets to assist in identifying suspects involved in CSAM circulation or online grooming networks. While technologically helpful, this use also raises concerns about reliability and potential misuse without legal oversight.[23]

AI plays a crucial role in digital evidence analytics. CSA investigations frequently involve chat histories, emails, social media communication, metadata, and device logs. AI tools can automatically sort, categorize, and reconstruct timelines from extensive datasets, helping investigators trace grooming behaviour and communication patterns efficiently.[24]

AI is increasingly used for:

- Automated DNA profiling and probabilistic genotyping
- Image/video enhancement in child exploitation material and CCTV recovery
- Facial comparison and pattern matching across datasets
- Digital evidence triage (chat logs, metadata, geolocation timelines)
- Voice and text analytics to identify grooming behaviour, coercion, and threatening communications
- AI assisted detection and classification of Child Sexual Abuse Material (CSAM) using image hashing and automated content recognition

5 Evidentiary Admissibility and Algorithmic Opacity

Under the Bharatiya Sakshya Adhiniyam, 2023, scientific and electronic evidence is admissible only when the court is satisfied about its authenticity, reliability, and proper certification. The statute recognizes expert opinion and electronic records as valid forms of proof, but it presumes that the expert can explain the basis of the conclusion in a manner open to judicial scrutiny and cross examination.[25] A further concern is whether the defence can meaningfully challenge such evidence. Fair trial principles depend on the opportunity to test the reliability of prosecution evidence through cross examination. If the methodology of the AI system is undisclosed due to intellectual property claims or technical complexity, the

[21] John M. Butler, Advanced Topics in Forensic DNA Typing: Interpretation (2015).

[22] Europol, AI and Policing: The Benefits and Challenges of Artificial Intelligence for Law Enforcement (Europol Innovation Lab Observatory Report 2024); Laura Sanchez, Cinthya Grajeda, Ibrahim Baggili & Cory Hall, *A Practitioner Survey Exploring the Value of Forensic Tools, AI, Filtering, & Safer Presentation for Investigating Child Sexual Abuse Material (CSAM)*, 29 Digit. Investigation S124 (2019).

[23] Clare Garvie, Alvaro Bedoya & Jonathan Frankle, The Perpetual Line-Up: Unregulated Police Face Recognition in America (Georgetown Law Center on Privacy & Technology 2016).

[24] Eoghan Casey, Digital Evidence and Computer Crime (3d ed. 2011).

[25] Bharatiya Sakshya Adhiniyam, 2023, §§ 39, 57, 61, 63 (expert opinion and electronic records).

defence is deprived of the ability to question error rates, bias, or data integrity. This weakens the adversarial process that underlies criminal justice.[26]

Further, courts traditionally assess scientific evidence by examining validation studies, peer acceptance, and known error margins. In the case of AI tools, questions arise as to whether error rates, training datasets, validation protocols, and audit records are disclosed to the court. Without such disclosure, AI derived forensic conclusions risk being accepted on blind technological trust rather than proven reliability.[27]

6 IMPLEMENTATION GAPS: INFRASTRUCTURE AND TRAINING

Despite the growing recognition of Artificial Intelligence (AI) as a valuable tool in forensic investigations, its effective implementation in child sexual abuse (CSA) cases remains constrained by significant institutional and operational deficiencies. The availability of AI enabled forensic infrastructure is uneven across Indian States, with many forensic laboratories lacking the technological capacity, specialised software, and computational resources required to deploy advanced AI applications. These disparities contribute to delays in forensic examination and limit the consistent use of AI assisted techniques in criminal investigations.[28]

The successful integration of AI into forensic practice also depends on the competence of the professionals responsible for its use. However, specialised training on AI assisted forensic technologies remains limited among investigating officers, forensic experts, prosecutors, and members of the judiciary. In the absence of adequate technical knowledge, there is a risk that AI generated findings may either be accepted without sufficient scrutiny or rejected due to a lack of understanding of their scientific basis.[29] This challenge is further compounded by the absence of standard operating procedures governing the validation, verification, documentation, and audit of AI assisted forensic tools. Without uniform protocols, the reliability and reproducibility of AI generated evidence may be questioned during judicial proceedings.

Another significant concern is the increasing dependence on proprietary AI software developed by private entities. Many of these systems operate through closed algorithms that are neither independently audited nor subject to mandatory regulatory certification. The absence of statutory standards for algorithmic transparency, software accreditation, and forensic validation creates uncertainty regarding the reliability and admissibility of AI assisted evidence.[30] Consequently, the implementation gap extends beyond technological limitations and reflects a broader disconnect between India's progressive legal

---

[26] Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018).

[27] PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY, FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS (2016); P. Jonathon Phillips et al., *Four Principles of Explainable Artificial Intelligence* (National Institute of Standards and Technology, NISTIR 8312, 2021).

[28] Ministry of Home Affairs, Government of India, Reply to Lok Sabha Unstarred Question No. 3452, *Infrastructure in Forensic Laboratories* (Dec. 17, 2024); Press Information Bureau, Government of India, *National Forensic Infrastructure Enhancement Scheme* (July 31, 2024).

[29] DEPARTMENT-RELATED PARLIAMENTARY STANDING COMMITTEE ON HOME AFFAIRS, RAJYA SABHA, REPORT NO. 254, CYBER CRIME: RAMIFICATIONS, PROTECTION AND PREVENTION ¶ 4.1.16 (Aug. 20, 2025); Laura Sanchez, Cinthya Grajeda, Ibrahim Baggili & Cory Hall, *A Practitioner Survey Exploring the Value of Forensic Tools, AI, Filtering, & Safer Presentation for Investigating Child Sexual Abuse Material (CSAM)*, 29 DIGIT. INVESTIGATION S124 (2019); UNITED NATIONS OFFICE ON DRUGS AND CRIME (UNODC), EDUCATION FOR JUSTICE (E4J) UNIVERSITY MODULE SERIES: CYBERCRIME, MODULE 4: INTRODUCTION TO DIGITAL FORENSICS (2019).

[30] UNESCO, RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE (2021); NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST), ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0) (2023); EUROPEAN COMMISSION, ETHICS GUIDELINES FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (High Level Expert Group on AI 2019).

framework for child protection and the institutional preparedness required to responsibly integrate AI into forensic investigations.

7 TOWARD A CHILD CENTRIC AI FORENSIC FRAMEWORK

The paper proposes:

1. Statutory guidelines for AI use in CSA forensics

There is a need for statutory or delegated guidelines specifically regulating the use of AI in CSA forensics. These guidelines should define permissible uses, validation requirements, documentation standards, and limitations to ensure that AI tools operate within the protective mandate of child friendly investigation.[31]

2. Accreditation and audit of forensic AI tools

Second, accreditation and periodic audit of AI forensic tools must be made mandatory. Independent forensic authorities should certify software accuracy, test error rates, and verify reliability before such tools are used in investigations. This prevents unverified proprietary systems from influencing criminal trials.[32]

3. Disclosure norms for algorithmic reliability in court

Third, courts must require disclosure norms when AI derived evidence is produced. Details regarding algorithmic methodology, validation studies, and known limitations must be placed on record to allow meaningful cross examination under the Bharatiya Sakshya Adhiniyam, 2023.[33]

4. Child data minimization and deletion standards

Fourth, strict child data minimization and deletion standards must be enforced in line with the Digital Personal Data Protection Act, 2023. AI systems must not retain or reuse child forensic data beyond the purpose of investigation and trial. Encrypted storage, access control, and mandatory deletion timelines are essential.[34]

5. Specialized training for justice stakeholders

Fifth, specialized training should be provided to police officers, forensic experts, prosecutors, and judges. Stakeholders must understand both the capabilities and limitations of AI tools to prevent blind reliance on technological outputs.[35]

6. Judicial checklists before admitting AI derived evidence

Sixth, a judicial admissibility checklist should be evolved for courts before accepting AI derived forensic conclusions. This checklist may include verification of tool accreditation, disclosure of error rates,

---

[31] *See* Ministry of Home Affairs (Women Safety Division), Government of India, Advisory on Mandatory Action by Police in Cases of Crime against Women, F. No. 15011/190/2020-SC/ST-W, ¶ 2(vi) (Oct. 2020) (noting the guidelines of the Directorate of Forensic Science Services for the collection, preservation and transportation of forensic evidence in sexual assault cases).

[32] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, ISO/IEC 17025:2017, GENERAL REQUIREMENTS FOR THE COMPETENCE OF TESTING AND CALIBRATION LABORATORIES (2017); FORENSIC SCIENCE REGULATOR (U.K.), CODE OF PRACTICE, Version 1 (Mar. 2023).

[33] Bharatiya Sakshya Adhiniyam, 2023, §§ 39, 57, 61.

[34] Digital Personal Data Protection Act, 2023, §§ 4, 5, 9.

[35] DEPARTMENT-RELATED PARLIAMENTARY STANDING COMMITTEE ON HOME AFFAIRS, RAJYA SABHA, REPORT NO. 254, CYBER CRIME: RAMIFICATIONS, PROTECTION AND PREVENTION ¶ 4.1.16 (Aug. 20, 2025).

expert explanation of methodology, and confirmation of data protection compliance. Such a checklist would operationalize fairness in practice.[36]

8 CONCLUSION

AI enhanced forensics holds transformative promise for POCSO investigations. Yet, without a child centric legal framework ensuring transparency, accountability, and data protection, these tools risk undermining fair trial standards and child rights.

AI integration into forensic processes under the Protection of Children from Sexual Offences Act, 2012 offers transformative potential for investigating child sexual abuse with greater speed, precision, and scientific reliability. Tools such as facial recognition, digital pattern analysis, age estimation, and intelligent evidence sorting can reduce investigative delays and human error while strengthening evidentiary chains. In a system where forensic lapses often weaken prosecution, AI enabled methods can significantly improve the quality and timeliness of evidence collection and analysis.

A child centric regulatory framework is therefore essential to govern AI use in POCSO investigations, ensuring data minimization, privacy protection, algorithmic accountability, and judicial oversight. Embedding transparency, accreditation of forensic AI tools, and specialized training for justice stakeholders can convert technological capacity into genuine social justice. The path to Viksit Bharat 2047 lies not merely in adopting AI, but in governing it wisely for the most vulnerable.

* * *

[36] Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473; Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1.

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.7

CHAPTER 7

Artificial Abuse: A Socio-Legal Analysis on the Effect of Usage of Generative Artificial Intelligence Against Vulnerable Populations

SHARON THOMAS* AND SOORYA NARAYANAN**

The rapid increase in the reliance on Artificial Intelligence in everyday life has resulted in the creation of a new kind of gendered harm, in the form of exploitative materials created using Generative Artificial Intelligence (A.I.). This has led to an expansion of existing gaps between technological development and legal protection. This paper examines the intersection of images produced using A.I. tools and legal remedies available for the same, with a specific focus on how these tools are used as another medium to harass women. It analyses how A.I. exacerbates structural inequalities and highlights the critical need for reforms in the current Indian criminal justice system. Several forms of online harassment that are already faced by individuals such as stalking, image alteration through creation of deepfakes, sexual harassment, nonconsensual recording etc. are adding to the psychological and sexual abuse against women, making their commission easier and their prosecution harder. We also explore the technological gaps in the creation of these A.I. tools, especially the gender bias in training A.I., and its effect on the vulnerable populations of India where caste, class, religion, and disability intensify the vulnerability for women. The psychological and economic impact on the victims including trauma, digital withdrawal, job loss, reputational harm, and barriers to online participation are understood through case studies. Finally, the paper examines the legal framework addressing the digital abuse against women by comparing international legislations like the European Union A.I. Act and United Nations guidelines with the Indian legislations and identifies gaps in the Bharatiya Nyaya Sanhita, Information Technology Act, and Digital Personal Data Protection Act. The paper concludes by proposing recommendations to make AI governance gender-responsive and intersectional.

Keywords: Artificial Intelligence, Generative AI, Deepfakes, Gender-based Violence, Discrimination.

1 INTRODUCTION

The world is entering into the era of Artificial Intelligence (AI), where technology has been rapidly integrated into various aspects of everyday life and systems. The scale of development and training of these AI models, especially Generative AI models, has moved faster than regulators can address. This has led to the creation of statutory and regulatory vacuums which have allowed the harm caused by AI systems to vulnerable populations to slip through the regimes. These vulnerable populations include women, children, individuals belonging to the LGBTQIA+ community, individuals belonging to poor communities and more. Through this paper, we explore the disproportionate harm caused to them by AI systems, the systematic reinforcement of discrimination due to the way these models are created and trained, real-life examples of how these tools have actively harmed individuals, existing statutory frameworks and their

* Student at VIT School of Law, Vellore Institute of Technology, Chennai, Tamil Nadu, India.
** Student at VIT School of Law, Vellore Institute of Technology, Chennai, Tamil Nadu, India.

scope and final recommendations to create a safer framework to utilize these tools in a way that reduces the harm to these populations.

2 UNDERSTANDING AI AS A TOOL FOR GENDERED VIOLENCE

Artificial Intelligence can be understood as the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. Generative AI technology is a subset of the same and refers to technology that is trained on large amounts of data that can generate content in the form of text, audio, video etc. in response to 'prompts' given to them. Rapid technological advances in the field of AI are creating unprecedented risks to facilitate violence against women. Technology-facilitated violence against women and girls (TF VAWG) is comparable to AI-assisted and AI-generated digital violence, but the generative AI models have greater access to data that make them more harmful and realistic. The humongous scale and unpredictability of AI models enable them to cause significantly more harm. The added layer of anonymity of the perpetrators makes access to justice even harder.

Common forms of online TF VAWG are mostly exacerbated by the usage of AI tools, which increases the harm that is caused by them. The volume of media that is created through Generative AI tools has obscured the distinction between genuine information and generated information, which is utilized to increase malicious targeted disinformation to a greater audience. The promulgation of misinformation is also easier as the AI models are trained on datasets that are inherently biased against women, reinforcing gender stereotypes. The non-consensual sharing of personal information is made much easier through AI tools, which can identify vulnerabilities and exploit them when interacting with female victims.

Gender inequality in society is reflected completely in these AI systems and is embedded in every part of their life cycle, from the people developing the AI to who accesses it and finally to who regulates it. A study conducted in 2021 analyzing 133 AI systems across various industries found that about 44.2% (59) of them showed gender bias, and 25.7% (34) exhibited both gender and racial bias[1]. When the foundation of AI runs entirely on data and information that is given to it, the systems are built to perpetuate harm from their inception. The underrepresentation of women systematically in AI research and development influences the data used, priorities undertaken and evaluation of the systems. The structural challenges faced by women, which include lack of digital access, limited financial ability, unequal education, care responsibilities and lack of relevant training, restrict their ability to even use these AI systems which have quickly become part of everyday life. Women risk becoming passive users or completely excluded from AI opportunities, reinforcing the digital inequalities that facilitate Technology-facilitated Gender-based Violence (TFGBV). Without safeguards, these systems risk deepening inequality[2].

[1] G. Smith & I. Rustagi, *When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity*, STAN. SOC. INNOVATION REV. (Mar. 31, 2021).

[2] Jinha Kim & Byeogjo Kong, *Bridging the Digital Divide: Harnessing Artificial Intelligence for Gender Equality in Asia and the Pacific* (ADB Brief, Oct. 2025).

3 FORMS OF AI ABUSE AGAINST WOMEN

3.1 Deepfakes and Non-Consensual Imagery

A deepfake is a digital image, video, voice, etc. that has been convincingly generated or altered to misrepresent someone as doing or saying something that was not actually done or said. Deepfakes are primarily made by GANs (Generative Adversarial Networks) which is an artificial intelligence system where two neural networks compete against one another to form hyper-realistic synthetic data[3]. A GAN system works through multiple competitive loops where the outputs which create synthetic images are repeated till it achieves the most accuracy that a common man would not be able to differentiate from a real image[4]. However, the increase in access to the use of AI has made it easy for non-tech users who do not have special expertise in computer or AI systems to be able to generate these hyper-realistic images within seconds, with a much larger number of people now using these tools for this purpose.

Another form of TFGBV is non-consensual Synthetic Intimate Imagery (NCSII), commonly known as revenge porn, which reveals a structural form of electronic sexual violence that does not comply with traditional sexual assault laws that exist making it difficult to handle such cases[5]. Historically, most non-consensual sharing of sexually explicit images happens within intimate relationships where the photos could've been taken with consent but shared against the will of the victim. With the rise of Generative AI and easier access to synthetic images, the earlier requirement of physically capturing the photo isn't necessary, and almost anyone can fall victim to this form of gendered violence. These deepfakes can be created with just an innocent headshot of a person, and any other images which can be easily found online.

The widespread access to deepfake technology affects public personalities and ordinary people alike. One of the earliest cases in India which made people aware of the dangers of AI technology was when the actress Rashmika Mandanna's face was digitally superimposed onto an explicit video of another woman[6]. The incident which occurred in November of 2023, soon after Gen AI was made accessible to the public, sparked a huge debate on the dangers of it, and led law enforcement to act against the perpetrator under the Information Technology Act (IT Act) and the Indian Penal Code (IPC). In January 2024, the 24-year-old accused was arrested for online sexual harassment. These cases do not happen only against high-profile public figures. Schools, universities, and workplaces often have cases where male workers use non-consensual synthetic imagery as a weapon to humiliate, silence, or blackmail their female peers[7]. These images spread quickly throughout the internet, ruining the reputation of the victims and affecting their everyday personal and professional lives, both online and offline.

---

[3] Preeti, M. Kumar & H.K. Sharma, *A GAN-Based Model of Deepfake Detection in Social Media*, 218 PROCEDIA COMPUT. SCI. 2153–62 (2023).

[4] B.N. Ramanizan, *Generative Adversarial Networks (GANs): A Comprehensive Review of Architectures, Training Challenges, and Advancements*, 7 J. ARTIF. INTELL. & CAPSULE NETWORKS 304–19 (2025).

[5] Preeti, M. Kumar & H.K. Sharma, *A GAN-Based Model of Deepfake Detection in Social Media*, 218 PROCEDIA COMPUT. SCI. 2153–62 (2023).

[6] S. Mankermi, *Etimes Explainer: The Truth Behind Rashmika Mandanna's Deepfake Video, Decoded with Violation of Privacy Rules, Legal Implications and Technological Drawbacks*, HINDI MOVIE NEWS (Nov. 8, 2023), etimes.in.

[7] A. Karagianni & M. Doh, *A Feminist Legal Analysis of Non-Consensual Sexualized Deepfakes: Contextualizing Its Impact as AI-Generated Image-Based Violence Under EU Law*, PORN STUD. 1–18 (2024).

3.2 AI-Enabled Stalking and Surveillance

Modern stalking tools are being developed by utilising predictive AI where a person's routines can be tracked through location data and analyzing past movements to forecast possible future outcomes. Perpetrators can use cheap tracking devices which they can attach to the victim's personal belongings or vehicles and connect them to AI models which analyse the data and provide real life updates on the location along with future predictions[8]. This allows stalkers to predict when their victims would be most vulnerable, to physically confront or harm them.

Modern AI has also easily taken advantage of facial recognition technology. For example, PimEyes is one of several facial-search services reported to index images scraped, without the consent of the people pictured, from public websites and social media where even one picture of the person can reveal their full identity along with personal details[9]. It isn't necessary for there to be a picture of the woman available on the internet as an individual can even capture an image of a woman in public without her knowledge and use that picture to track her down. Such searches can reveal employment histories, locations, family members or regularly visited public areas of the victim, threatening their safety and privacy.

3.3 AI in Intimate Partner Violence

Domestic violence advocates have begun to document regular cases where abusive partners use security cameras, smart door locks and other AI home technology to automatically lock doors trapping their partner with them or manipulate security alarms to control and isolate the victim[10]. These physical forms of interruption with AI tools can lead to coercive physical and emotional forms of control, and abuse, with very minimal force. This method of tracking a partner's domestic life leads to a severe lack of privacy leading to psychological distress.

Abusers utilize smart devices along with AI technology to manipulate their victim's perception of reality, easily manipulating and dominating them[11]. They can alter the victim's digital records of plans in their calendar, delete home security camera footage, and use AI voice modification to manipulate conversations which creates an extremely stressful and confusing environment for the victim to navigate. These forms of manipulation are easy to commit but hard to prove, reducing the already slim chances for the victim to get justice. This form of AI tracking and surveillance further increases the chances for physical and sexual violence since the abuser has full knowledge of their daily routines and can discard the evidence. Constant surveillance also reduces the likelihood for the victims to plan an escape route as all their movements are tracked. Victims often fear that their abusers could have non-consensually recorded or generated sexually explicit videos or images that they could publicize to shame and control them even further.

---

[8] Nameeta Rana, *Artificial Intelligence, Cyber Stalking, and the Future of Digital Privacy: A Legal Perspective*, 6 INT'L J. FOR MULTIDISCIPLINARY RSCH. (2024).

[9] J. Wakefield, *PimEyes Facial Recognition Website "Could Be Used by Stalkers"*, BBC NEWS (June 11, 2020), https://www.bbc.com/news/technology-53007510.

[10] A. Brown, D. Harkin & L.M. Tanczer, *Safeguarding the "Internet of Things" for Victim-Survivors of Domestic and Family Violence: Anticipating Exploitative Use and Encouraging Safety-by-Design*, 31 VIOLENCE AGAINST WOMEN 1039–62 (2024).

[11] Peter Novitzky, *A Systematic Review of Ethical Challenges and Opportunities of Addressing Domestic Violence with AI-Technologies and Online Tools*, 9 HELIYON (2023).

3.4 Non-Consensual Recording through AI Devices

The creation and distribution of AI hardware products such as camera-equipped smart glasses, including Meta's AI glasses, whose small cameras and microphones may go unnoticed by others, can enable people to record others in public without their consent. The increasing use of Meta AI glasses has raised concerns regarding violation of privacy and bodily autonomy[12]. Because such devices can look like ordinary accessories, they may be misused to violate someone's privacy. This is a violation of Article 21 of the Indian Constitution[13], and section 77 of the Bharatiya Nyaya Sanhita (BNS)[14] which handles voyeurism, that is, the non-consensual recording or sharing of private videos or pictures taken without consent. Despite this, due to the elusive nature of the AI devices, it is highly improbable that the perpetrators are caught. The concern for safety escalates with devices that are attached to real-time facial recognition software and LLMs. Researchers have found that these devices allow the perpetrators to access all the private information about a person with the help of AI[15].

4 Social and Intersectional Issues with AI Abuse

4.1 Gender Bias and Structural Discrimination in AI Systems

AI is trained on available data on the internet, including historical statistics and structures. When LLMs consume vast amounts of data from the internet, they also collect and learn the biased discriminatory systems and practices followed historically, which leads to algorithmic bias. AI models[16] thus end up learning sexism, racism, classism, transphobia, and other forms of discriminatory information. For example, when an AI tool is told to generate a person in a leadership position, it would generate images of men. In contrast, when it is told to generate a person in a caregiving job, it would generate a woman[17]. This bias affects job opportunities for women since AI data would show preference for male workers over female workers in most jobs, aggravating the existing issues of gender inequality in workplaces.

4.2 Intersectionality

To understand intersectional bias within AI, it is important to first understand the intersectional framework where overlapping systems of discrimination cause a layered form of oppression. Kimberlé Crenshaw, a legal scholar and an intersectional feminist, showed how structures of marginalization overlap, making individual experiences different. She highlighted[18] how a white woman and a black woman would have different forms of discrimination that they must overcome, and how Black women can face discrimination on the combined grounds of race and sex that neither a gender-only nor a race-only analysis captures. Similarly, a woman's sexuality, class, caste, etc. can intensify the form of violence, harassment,

---

[12] Suraksha P., *Meta AI Glasses Pilot Faces Privacy Scrutiny*, ECON. TIMES (June 20, 2026).
[13] INDIA CONST. art. 21.
[14] Bharatiya Nyaya Sanhita, 2023, § 77.
[15] Kaira Software, *AI Privacy in Crisis: 5 Scandals That Expose the Dark Side of Artificial Intelligence*, KAIRA SOFTWARE, https://kairasoftware.com/blog/ai-privacy-in-crisis-5-scandals-that-expose-the-dark-side-of-artificial-intelligence (last visited Sept. 22, 2026).
[16] Nivedhaa N, *A Comprehensive Review of AI's Dependence on Data*, 1 INT'L J. ARTIF. INTELL. & DATA SCI. (IJADS) 1–11 (2024).
[17] A. Bhardwaj, *AI Is Already Rewriting Reality, but Getting Women Wrong, Says UN Women*, BUS. STANDARD (June 23, 2026), https://www.business-standard.com/technology/artificial-intelligence/ai-is-already-rewriting-reality-but-getting-women-wrong-says-un-women-126062300887_1.html.
[18] K. Crenshaw, *Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color*, 43 STAN. L. REV. 1241–99 (1991).

or discrimination she faces[19]. This is applicable to Artificial Intelligence, as it feeds off online data and learns the same biases, leading to marginalized women such as those from minority religions, women with disabilities, Dalit and Adivasi women, and women from the LGBTQIA+ community often facing targeted algorithmic attacks and abuse online.

4.3 Psychological and Economic Impacts

Survivors of non-consensual deepfakes and AI-surveillance based stalking face severe psychological trauma, similar to the clinical mental impact caused by physical sexual assault and stalking[20]. The realistic quality of modern AI and its generative images create severe emotional distress and anxiety among survivors of such attacks. Victims have reported that this form of AI crime has taken away their bodily autonomy and often get clinically depressed and develop Post Traumatic Stress Disorder[21]. To cope with this online abuse, victims often have no choice but to log off from all social media accounts and withdraw from public and professional areas to avoid hate, judgement, and scrutiny[22]. This withdrawal in turn leads to the removal of women from public media, and their voices being silenced and suppressed. Additionally, there are several economic impacts on women who face AI-harassment. Once their deepfakes become easily publicly accessible to anyone, employers may feel hesitant in letting the woman work in the company due to fear of public and moral perceptions, and may terminate her employment. Victims may end up losing their jobs over something they did not partake in, aggravating their trauma[23]. Further, in abusive relationships where the partner uses AI tools to monitor the woman's financial and banking tools, it becomes difficult for the victim to have economic independence. These structural issues risk widening the gender pay gap and make it more difficult for women in leadership positions to succeed.

5 LEGAL AND REGULATORY FRAMEWORKS

5.1 International Legislative Responses

The European Union's Artificial Intelligence Act or the EU AI Act (2024)[24] establishes a thorough and risk-based regulatory system that sorts AI usage into different and specific risk tiers based on the potential harm it would cause to society. It puts an explicit ban on any AI use that poses an unacceptable risk, such as extracting and scraping of facial images from CCTV footage or content from the internet to develop facial recognition databases, and biometric classification systems that deduce or infer a person's race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation. The EU AI Act enforces strict obligations on transparency for AI generated content to tackle synthetic manipulation.

The United Kingdom's Online Safety Act (2023)[25] acknowledges technology-based violence and abuse and includes criminal penalties for offenders. The Act makes the act of sharing non-consensual intimate media a punishable crime. It shuts down several legal loopholes regarding online harassment and

[19] Prince Kumar, *Intersections of Caste and Gender in Social Stratification*, SOCIOLOGY. INSTITUTE (2026).
[20] C. GOLDBERG, NOBODY'S VICTIM: FIGHTING HARASSMENT ONLINE AND OFF (2017).
[21] Muhammad Umar, *Digital Trauma: Deepfake Victimisation and AI-Generated Violence*, 13 LANCET PSYCHIATRY (n.d.).
[22] M. Ali, *Deepfakes and Victimology: Exploring the Impact of Digital Manipulation on Victims*, 8 SUBSTANTIVE JUST.: INT'L J.L. (2025).
[23] D. Bradley, *Deepfakes, Job Losses, Opaque Models: Exploring the Dark Side of AI*, TECH EXPLORE (2026).
[24] European Union's Artificial Intelligence Act or the EU AI Act (2024).
[25] Online Safety Act 2023 (UK).

abuse that allowed perpetrators to walk free without facing any sanctions. The Act also imposed a statutory duty of care for all digital media. Digital platforms are legally mandated to notice and remove non-consensual sexual content. The penalty for non-compliance is heavy, ensuring strong enforcement of the provisions.

The United States' legal responses to AI technology have been developed through federal and state level measures. The DEFIANCE Act – Disrupt Explicit Forged Images and Non-Consensual Edits Act[26], would establish a civil cause of action for victims. This act would grant the victims of explicit deepfakes the right and opportunity to sue the creator of the image, the people who distributed it, and anyone else who facilitates the crime for monetary damages. Several individual states like California, Texas, and New York have made criminal and civil statutes that punish non-consensual image creators and distributors.

International organizations have established certain ethical standards over the years for national guidance when drafting legislation on Artificial Intelligence. The UNESCO Recommendation on Ethics of AI (2021)[27] requests its member states to create special provisions and resources to reduce the gender biases in AI training databases and to make clear standards regarding accountability for gender-based violence or harassment through AI. The OECD Recommendation on Artificial Intelligence (2019)[28], furthermore, stresses human-centric values and transparency in the way AI is consumed and used. Although these international frameworks provide a good base for the way AI law is legally handled, their non-binding nature limits their usage across countries.

5.2 Indian Legal Framework

The basis of the laws addressing gender-bias in India is the Constitution. Article 14[29] of the Constitution guarantees equality in the eyes of law and provides the foundation to challenge and question systemic inequalities including gender-based discrimination or violence. Article 19(1)(a)[30] protects the freedom of speech and expression of individuals, which may be affected for victims of AI-based violence or non-consensual image creation as they are the ones who bear the consequences of such acts. The Supreme Court case of Justice K.S. Puttaswamy Vs. Union of India (2017)[31] declared the importance of personal dignity and rights over bodily autonomy as fundamentally protected under Article 21 of the Constitution[32]. Non-consensually morphing or generating an explicit image or video of someone directly infringes the ruling set by this court.

The Information Technology Act (2000), which is the primary Indian law for cybercrime, contains provisions for digital harassment. These provisions came into existence before the widespread usage of Generative AI and are therefore outdated in some scenarios. However, some provisions can still be relied upon such as Section 66E[33] of the act, which criminalizes intentional capture, publication, or transfer of

[26] Disrupt Explicit Forged Images and Non-Consensual Edits Act of 2025, S. 1837, 119th Cong. (passed by the US Senate on 13 January 2026; not yet passed by the House of Representatives).
[27] RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE, 41 C (2021).
[28] RECOMMENDATION OF THE COUNCIL ON ARTIFICIAL INTELLIGENCE (May 22, 2019).
[29] INDIA CONST. art. 14.
[30] *Id.* art. 19(1)(a).
[31] Justice K.S. Puttaswamy v. Union of India (2017).
[32] INDIA CONST. art. 21.
[33] The Information Technology Act, 2000, § 66E.

images taken of private body parts without consent. This provision can be used for non-consensual recording with Meta AI glasses but will not apply to generated images. Sections 67 and 67A punish publication and transfer of obscene or sexually explicit media[34]. These sections do not mention the capturing of images and therefore can be applied to AI generated images as well.

The Bharatiya Nyaya Sanhita (2023) updated several gender-based provisions in criminal law. Section 78 of the act penalizes physical and online stalking of a person[35]. Monitoring a woman's movements or electronic communication without her consent is punishable under this section. Section 79 punishes anyone who tries to insult the modesty of a woman through words, actions, and gestures[36]. This would include creating and sharing non-consensual explicit images. Section 356 of the act handles defamation, i.e., false actions or words that will cause damage to someone's reputation[37]. That section can be applied to cases where a woman's morphed or AI-generated images are created and shared without her consent, holding the offenders criminally liable.

The Digital Personal Data Protection Act (2023) requires owners of online websites to process the personal data of an individual only after receiving explicit consent. Yet, the act also contains exceptions that instantly weaken the effectiveness of the provision for non-consensual deepfake content. Under section 3(c)(ii) of the act, it exempts protection for personal data that is made publicly available by the individual themselves[38]. This means that women who post public images for professional or social purposes can have their images used by Gen-AI users with bad intent and still be left without protection.

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 makes it mandatory[39] for websites to take down non-consensual sexually explicit content within 24 hours of receiving a complaint. However, this is not followed in the strictest sense due to several restrictions, like the rapid increase in such content, low staff to handle such issues, and difficulty keeping track of them. Moreover, non-consensual digital content spreads rapidly among several different networks, websites, apps, etc. which may not always fall under India's legal jurisdiction.

6 WAY FORWARD

6.1 Ethics in Design and Technical Watermarking

To reduce the AI-assisted and enabled abuse that is committed against vulnerable communities, software developers must construct safety standards for the foundational structure of the AI systems and ensure that they are followed. These standards would mandate digital watermarks to accentuate AI-generated content[40], and the integration of cryptographic systems like the C2PA protocol into all generative platforms[41]. Cryptographic data lets web browsers, forensic tools, and content platforms verify whether media was artificially generated and automatically filters the content to ensure only intended recipients will

[34] *Id.* §§ 67, 67A.
[35] Bharatiya Nyaya Sanhita, 2023, § 78.
[36] *Id.* § 79.
[37] *Id.* § 356.
[38] Digital Personal Data Protection Act, 2023, § 3(c)(ii).
[39] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021.
[40] S. Tiwari, *Mandatory Watermarking and Labelling of AI-Generated Content* (2026).
[41] Lisa Braswick, *C2PA Explained – Content Credentials Guide*, AFIP (2026).

access it[42]. AI developers must also implement strict boundaries for the training models which makes it possible for them to reject prompts regarding non-consensual nude image generations or face-swapping without verification[43]. Incorporating such safety measures into the software models themselves would lower the possibility of digital harassment.

6.2 Liability of AI Platforms and Statutory Duty of Care

Relying only on individual reports to address harms is insufficient due to the high spread of non-consensual synthetic content on digital platforms. Social media regulations must establish a statutory duty of care on all social media platforms and hold them accountable for negligence if they fail to take down the non-consensual deepfakes or morphed images. Digital platforms must also implement automatic algorithms that will be able to identify, block, and report non-consensual imagery before they are uploaded. Statutory guidelines can also reduce mandatory takedown time frames from 24 hours to 2 hours for non-consensual sexually explicit content as they can spread quickly and cause irredeemable damage. Platforms that fail to follow these regulations should face legal disputes and must pay fines or penalties for damages.

6.3 Legislative Reform for India and Victim-Centric Approach

India must enact an act for AI safety and governance as it still does not have a specific framework for these tools. The act must focus specifically on regulating automated systems and protecting individuals from AI-enabled abuse. This act must define, criminalize, and punish non-consensual creation, possession, or distribution of sexually explicit content. The courts must ensure that a victim-centric approach is followed throughout the justice system to reduce their trauma and stress. The state should make the reporting process easier and more accessible for victims, where they can make one online report which must immediately trigger the takedown of the content across all online platforms and track down the perpetrator. Governments must also provide special services for the victims to help with their trauma by offering them psychological care free of cost, support groups, and safety services for anyone whose well-being was threatened after reporting. Special cyber-crime units with staff trained at understanding gender sensitivity and privacy rights and AI abuse are necessary for helping the victims recover faster.

7 Conclusion

The development of technology provides opportunities for the growth of society at large; however, this development must not be at the expense of those who are most vulnerable to the harm caused by these same technological developments. Understanding how Artificial Intelligence can be used as a tool to facilitate violence against marginalized communities ensures that further technological development is followed by appropriate regulations. By understanding the real-life harm caused by these virtual tools, Artificial Intelligence development can be done more holistically, by including all relevant stakeholders who are likely to be affected by its large-scale deployment across sectors. Despite the existence of multiple statutes that can be used to govern Artificial Intelligence in India, there are still significant statutory gaps

[42] S. Latha, *AI-Driven Identification of Cryptographic Algorithms from Encrypted Data*, Int'l J. Rsch. Tech. & Innovation (IJRTI) (2024).
[43] A. Trent, *Ethical AI Image Generation: Best Practices for Developers* (2026).

that necessitate the creation of a framework to regulate, supervise and govern Artificial Intelligence in the country.

* * *

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.8

CHAPTER 8

Cybercrime in the Metaverse: Identity, Jurisdiction and Digital Evidence

RUPALI MEHTA[*] AND ASHWINI YADAV[**]

The metaverse is emerging as a network of persistent, immersive and economically active digital environments in which people communicate through avatars, acquire virtual assets and generate continuous streams of behavioural and biometric data. The same architecture that produces presence and interoperability also enlarges the opportunity structure for crime. Identity theft may become avatar takeover or biometric mimicry; fraud may be executed through smart contracts and virtual goods; harassment may be embodied and spatial; and proceeds may move rapidly across wallets, platforms and states. This article examines three connected problems: attribution of conduct to a legally responsible person, selection and coordination of jurisdiction, and collection of reliable immersive digital evidence. Using doctrinal and comparative analysis, it evaluates the Indian framework comprising the Information Technology Act, 2000, Bharatiya Nyaya Sanhita, 2023, Bharatiya Nagarik Suraksha Sanhita, 2023 and Bharatiya Sakshya Adhiniyam, 2023, alongside emerging international cooperation models. Existing law can address much metaverse-enabled wrongdoing through technology-neutral offences, but important gaps remain where virtual identity, psychological injury, platform control and volatile multimodal evidence do not fit inherited legal assumptions. The article proposes a layered governance model based on functional identity assurance, harm-sensitive offence interpretation, effects-based jurisdiction, rapid preservation, standardized metaverse forensics and accountable platform cooperation. It argues that effective enforcement does not require universal real-name surveillance. It requires proportionate, auditable mechanisms that can connect serious conduct to accountable persons while preserving anonymity, privacy, due process and legitimate experimentation in immersive environments.

Keywords: metaverse; cybercrime; digital identity; avatar; jurisdiction; electronic evidence; virtual assets; India; platform governance.

1 INTRODUCTION

It is important to think of the metaverse not as a website or fully developed piece of tech, but as a socio-technical environment that is continually developing. It encompasses extended reality, persistent virtual worlds, avatars, artificial intelligence, spatial computing, payment systems and in some cases, blockchain-based assets. Users don't just read or post. They seem to share space and to communicate by voice and gestures; they play with objects, with transactions, building up reputations over time. These attributes can enhance social presence and economic engagement, while also introducing new challenges to familiar online risks and vulnerabilities. Inappropriate behaviour that had previously been expressed textually might now manifest as embodied intrusion. When an account is compromised, an avatar, inventory, social

[*] Research Scholar at Department of Law, Panjab University, Chandigarh, India.
[**] Research Scholar at NIILM University, Kaithal, Haryana, India.

history and access credentials are transferred in one. A fraud can consist of a fake avatar, an artificial voice, a bad smart contract and a wallet transfer in a matter of minutes[1].

It is not as if every metaverse crime is new. Common types of crime include theft, cheating, impersonation, stalking, intimidation, sexual abuse, obscenity, money laundering and unauthorized access. The challenge is that the metaverse reshapes the reality that those categories are usually based on. The victim is not seeing a real human but rather an avatar (actor) in front of him. The corresponding event could be broadcasted between a headset, a platform server, the cloud provider and the blockchain. The victim and suspect and/or service providers can be in different countries. Evidence is created at a fast pace and can be lost at the end of a session. Some harmful acts result in economic damage; others mainly in invasion of autonomy, dignity or psychological integrity. In this light, Qin, Wang and Hui (2025)[2] prioritise the idea of identity and unified enforcement in the governance of the metaverse, whereas Efremova and Russkevich (2025)[3] focus on anonymity, fuzzy legal status and procedural proof. Mokoena, Papadopoulou and Rodríguez (2023)[4] also highlight lack of evidence and conflict over territories as key challenges to prosecution.

In this article, three questions are raised. Firstly, how can the law be made to relate the activities of an avatar to a legally responsible person without sacrificing the anonymity that may be desirable? Second, what if the elements of a metaverse crime are spread across different states – which state has the authority to prescribe, investigate and adjudicate? Third, how can multimodal evidence, the meaning of which is dependent on a technical environment, be preserved, authenticated and presented to investigators? The questions are dependent on each other. To be able to perform jurisdiction, the point of attribution is necessary; to prove the point of attribution reliable evidence is necessary; to obtain reliable evidence in a timely manner, a lawful jurisdictional and cooperation mechanism is necessary.

It relies on doctrinal analysis of the Indian criminal laws and cybercrime laws, procedural and evidence laws with a comparative study of the literature and international instruments. It makes two assertions. First, under the current Indian legal framework, there is ample scope for tackling a significant number of offences against persons and property that can be facilitated by the use of the metaverse, including offences relating to unauthorised access or personation, cheating, stalking and intimidation, obscene or sexually explicit content, and laundering of criminal proceeds. Nevertheless, analogical application is problematic in the case of entirely virtual property, where purely psychological injury is the principal injury, where the conduct is materially produced by an AI system, or where the data on the platform is stored outside the country. Second, reforms should target functions, not a premature and comprehensive list of metaverse offenses. Technology-neutral definitions of harm, graded levels of identity assurance, fast data preservation, and standardized data capture and rights-protective cooperation would stand the test of time across virtual environments, rather than a distinct penal code for each.

---

[1] G.D. Ritterbusch & M.R. Teichmann, *Defining the Metaverse: A Systematic Literature Review*, 11 IEEE ACCESS 12368–77 (2023), doi: 10.1109/ACCESS.2023.3241809.

[2] H.X. Qin, Y. Wang & P. Hui, *Identity, Crimes, and Law Enforcement in the Metaverse*, 12 HUMAN. & SOC. SCI. COMMC'NS 194 (2025), https://doi.org/10.1057/s41599-024-04266-w.

[3] M.A. Efremova & E.A. Russkevich, *Criminal-Legal Issues of Countering Crime in the Metaverse: Current State and Prospects of Development*, 3 J. DIGIT. TECH. & L. 187–202 (2025), https://doi.org/10.21202/jdtl.2025.8.

[4] Thabo Mokoena, Eleni Papadopoulou & Camila Rodríguez, *Cybercrime Prosecution in the Metaverse: Evidentiary and Jurisdictional Challenges*, 2 LEGAL STUD. DIGIT. AGE 53 (2023).

2 METAVERSE AS A UNIQUE CRIME ENVIRONMENT

There is no universally agreed-upon definition of the metaverse. A useful definition, for purposes of the criminal law, should seek to focus on characteristics that affect opportunities, harm or proof. The metaverse can therefore be defined as a set of digitally mediated environments that can be persistent or recurring, and that provide a sense of spatial presence, synchronous interaction via avatars or agents, creation or possession and exchange of digital objects, and extensive processing of contextual, behavioural or biometric data. The persistence does not imply that all objects must persist forever, and the decentralization is not mandatory. A commercial virtual reality platform can be centrally controlled, while assets and identity credentials can be issued on blockchains and by independent providers[5].

There are four features which stand out in the environment. First, conduct is endowed with spatial and affective attributes. An intimidating avatar can come closer, accompany, surround or emulate unwelcome contact. Sensory feedback can be provided with haptic devices. According to González-Tapia (2023)[6], virtual emotions are of legal relevance because experiences experienced through technology can be real fear, humiliation and traumatic experiences. Not touching shouldn't be equated with no damage, especially if it is legally cognizable. Meanwhile, the criminalisation cannot be linked to all unpleasant virtual interactions, but only to deeds of culpability and objectively serious interference.

Second, the value of the identity, and also of property, increases thanks to persistence and portability. An avatar can have a reputation, employment access, creative output and purchased 'wearables' and social relationships. The ability to exchange virtual land, tokens and in-world items for money, and the scarcity of these items by platform design, also add to their value. Cheong (2022)[7] points out that there are questions in the areas of identity, property, intellectual property, and liability regarding avatars. A pure contract approach – where each asset is a revocable platform licence – may not capture the true reliance and economic loss. On the other hand, making all game objects property for all criminal-law purposes might overcriminalize breaches of the rules within fictional worlds. The parties' reasonable expectations, monetary convertibility, scarcity, and transferability and control should all be analyzed from a legal perspective.

Third, it's a data-intensive environment. Headsets and controllers can capture gaze, voice, facial expression, hand movement, body posture, and room geometry and the proximity to others. This information can be used to recognize users, detect disabilities or emotions, and even recreate actions with a high degree of accuracy. It is both proof and sensitive personal information. Based on this, McStay (2023)[8] characterizes the metaverse as a surveillant physics, as it is necessary to measure bodies and spaces to render the experience. At the very heart of the resultant paradox is the notion that the most probative record can also be the most intrusive.

Fourthly, the metaverse is a phenomenon of public and private governance. Platform operators write community rules, set safety boundaries, authenticate accounts, moderate behavior and control logs. They can investigate and sanction quicker than states, but with different aims, methods and transparency than criminal justice. Decentralization is not the same as having no control: decentralized autonomous

[5] T.A. Jaber, *Security Risks of the Metaverse World*, 16 INT. J. INTERACT. MOB. TECHNOL. 4–14 (2022).
[6] M.I. González-Tapia, *Virtual Emotions and Criminal Law*, 14 FRONTIERS PSYCH. 1260425 (2023).
[7] B.C. Cheong, *Avatars in the Metaverse: Potential Legal Issues and Remedies*, 3 INT'L CYBERSECURITY L. REV. 467–94 (2022), https://doi.org/10.1365/s43439-022-00056-9.
[8] A. MCSTAY, AUTOMATING EMPATHY: DECODING TECHNOLOGIES THAT GAUGE INTIMATE LIFE 304 (2023).

organizations and smart contracts can distribute control. According to Zwitter and Hazenberg (2021)[9], technology design itself has a normative power. The state should therefore walk hand in hand with the platforms, rather than delegating the definition of crime and coercive power and the fair procedural rules to the platforms' terms of service.

These features allow a progression of misbehaviors. On one extreme are the standard cybercrimes done in a new way: grabbing credentials, sending out malware and phishing and hacking accounts. Then there are technology enhanced crimes in the middle: deepfake personation, coordinated stalking, extortion with intimate avatar recordings, fraud in virtual asset markets, and virtual wallet layers. On the opposite side of the spectrum are harms that are dependent on the environment, such as simulated sexual contact without consent, manipulation of a child's avatar, or taking something that doesn't exist in reality and therefore has no value. But, as Marshall and Tompsett (2024)[10] rightly warn, the metaverse is not a completely new frontier. But continuity of offence labels should not be confused for changes in scale, intensity and evidentiary structure.

3 DIGITAL IDENTITY, DIGITAL ATTRIBUTION AND CRIMINAL RESPONSIBILITY

3.1 The Layered Identity Problem

In the metaverse, there is multiple layering to identity. A natural person could control multiple avatars, multiple people could share one account, one avatar could be controlled by an organization and an artificial intelligence agent could act with varying degrees of human direction. The presentation layer is using the display name. The layers below include device identifiers, IP records, wallet addresses, payment details, biometric patterns and platform logs, information that provides the necessary credentials for an account. Investigators need to navigate the layers carefully to be able to make attributions.

As a rule, an avatar is not a unique legal entity. The rights, duties, assets and responsibility must be allocated to legal persons, however visually autonomous or conversationally sophisticated the avatar may appear. But there should exist a specialised legitimate interest in the integrity of avatars. Even if the avatar does not have a human-like personality, privacy, reputation, property and personal autonomy can be undermined by unauthorised uptake, misleading reproduction or malicious modification. Mitrushchenkova (2022)[11] presents the personal identity in the metaverse as a site of expressive freedom, but also one of vulnerability. The rightful approach is to safeguard the human or organisation portrayed as the avatar and treat the avatar and credentials as evidence and, if appropriate, as proprietary objects.

There are valid reasons for anonymity. It enables social engagement, exploring disabilities, protection for vulnerable people and creative role play. A real name system would produce databases that would be loved by criminals, would allow for profiling, and provide a chilling effect on legitimate use. But full practical non-traceability can cover repeat offenders. A functional or tiered identity assurance is a proportionate model. Low risk social exploration can be pseudonymous. There is a possibility to provide more robust assurances for higher-risk functions such as large transfers, access to children, professional

[9] A. Zwitter & J. Hazenberg, *Cyberspace, Blockchain, Governance: How Technology Implies Normative Power and Regulation*, *in* BLOCKCHAIN, LAW AND GOVERNANCE 87–97 (B. Cappiello & G. Carullo eds., 2021), https://doi.org/10.1007/978-3-030-52722-8_6.

[10] A.M. Marshall & B.C. Tompsett, *The Metaverse—Not a New Frontier for Crime*, 6 WIREs FORENSIC SCI. e1505 (2024), https://doi.org/10.1002/wfs2.1505.

[11] A.N. Mitrushchenkova, *Personal Identity in the Metaverse: Challenges and Risks*, 9 KUTAFIN L. REV. 793–817 (2022), https://doi.org/10.17803/2313-5395.2022.4.22.793-817.

services, or operating regulated markets. The platform can validate an attribute or the platform could store an encrypted identity link without using a civil name. Disclosure is to take place through the course of a lawful and reviewable process.

3.2 Impersonation, Deepfake and Biometric Mimicry

Metaverse impersonation goes beyond copied usernames. Generative AI can duplicate voice, face, gesture and conversational style. A person trying to defraud may show up in a familiar virtual setting and instruct an employee to move assets to him or her. Biometric information obtained from motion sensors can be played back or used to deduce the identity. While Yang et al. (2023)[12] suggest frameworks for traceable authentication, not all authentication is traceable, as technical systems can be hacked and credentials can be shared, and classifiers of biometric data can be wrong.

There are multiple ways to enter into Indian law. Section 66C of the Information Technology Act, 2000 (IT Act)[13] covers the criminal use of another person's electronic signature, password or unique identification feature and section 66D of the IT Act covers cheating by personation using a communication device or computer resource. Unauthorized access, copying, introduction of contaminants and related conduct may lead to the imposition of sections 43[14] and 66[15], as long as the aforementioned statutory mental element is fulfilled. The Bharatiya Nyaya Sanhita, 2023 (BNS) includes general provisions of cheating and cheating by personation along with provisions relating to forgery which may be applicable to electronic records. They cover many deepfake-enabled frauds that might take the form of an avatar, and they are technology neutral enough to be applicable.

Gaps nevertheless remain. The specific wording in section 66C would apply best if a defined credential or a unique identification feature is misused. A convincing imitation can be created from publicly available voice and gesture information without the need to steal the credentials. Typically, personation offences will need to involve a degree of cheating, deception, or a similar result, and non-financial identity abuse could be covered by the privacy, defamation and harassment offences. Digital identity interference should be defined in functional terms, namely whether the interference has been intentional, unauthorized; whether it has caused or created a substantial risk of economic loss, reputational injury, sexual exploitation, unlawful access or risk of significant psychological harm. Include defenses where there is no prohibited deception, such as satire, art, research, and authorized simulation.

A person's responsibility for a crime remains the basis of the criminal law. If an AI-controlled avatar causes harm to another user, then the responsibility for the damage needs to be determined based on the connection between the human and the AI. No one should avoid facing the consequences for telling an agent to harass or defraud when that is what the agent was instructed to do, even if the last words were automatic or the last motion was made by the agent. It is the responsibility of a developer or deployer where an offence recognises recklessness, negligence, abetment and conspiracy with elements established. There is, however, no substitute for mens rea beyond simply having a system.

The evidentiary inquiry should clarify the agency chain that leads from the choice of objective through completion of the data, setup of the protective measures, approval of the deployment, keeping of the

[12] K. Yang, Z. Zhang, Y. Tian & J. Ma, *A Secure Authentication Framework to Guarantee the Traceability of Avatars in Metaverse*, 18 IEEE TRANSACTIONS ON INFO. FORENSICS & SEC. 3817–32 (2023), https://doi.org/10.1109/TIFS.2023.3288689.
[13] The Information Technology Act, 2000.
[14] *Id.* § 43.
[15] *Id.* § 66.

override power for its own uses and benefit, and the anticipated results. Logs of prompts, model versions and tool calls may be crucial. The risk is hindsight liability – the liability that one can automatically assume just because a complex system caused harm. Multiple actors present challenges when it comes to responsibility for AI as noted by Padovan, Martins and Reed (2023)[16]. There is a need to differentiate between whether the fault is committed, using criminal law to punish if it is, and whether the risks are sufficiently serious to warrant civil liability, regulatory measures and action on the platform to remediate.

3.3 Children's Experiences and the Needs of Vulnerable Users

Age deception, grooming, simulated sexual activity and the collection of sensitive data are specific dangers for children. Embodied communication can be more persuasive and thus more immediate in the area of grooming. While age assurance might be accepted for adults, having to take government identity papers from all children may pose another security risk. Use of privacy preserving age estimation or verified age tokens, if possible, including parental tools, child-friendly reporting, and personal boundaries and limits on unnecessary adult-child contact should be encouraged[17].

Transmission or publication within the meaning of the IT Act (including those that could be considered sexually explicit material or material that depicts children) may occur, but not every live embodied misconduct is necessarily a conventionally conceived image or a file. The legal aspect requires focussing on solicitation, grooming, intentional exposure and serious non-consensual simulation and at the same time maintaining the principle of legality. Platform rules may be more stringent than the criminal law as long as they are both known and subject to appeal in enforcement.

4 THE SUBSTANTIVE OFFENCES AND INDIAN LEGAL FRAMEWORK

4.1 Fraud, Unauthorized Access and Virtual Property

Social engineering can be paired with programmable transactions in the metaverse. A user can be tricked into linking a wallet, accepting a malicious smart contract, buying a fake virtual asset, and/or giving too much information to an avatar impersonating a support staff member. Both the IT Act and BNS will offer a practical combination of responses to an unauthorized computer related activity: the IT Act will permit prosecution for unauthorized computer-related activity, and deception and dishonest inducement can be prosecuted as cheating. Anti-money-laundering law applies if the statutory conditions are satisfied in cases where the proceeds are converted or layered using crypto-assets.

Theft of virtual property is more difficult. The traditional approach to theft has traditionally been associated with movable property and the actual taking of the property, but in some cases, a copy of a data source may be made without the original being removed. But for some virtual assets, in practice the control of a token or an account is transferred, and the victim loses the means of using or selling the asset. Strikwerda (2012)[18] suggests that, besides physicality, the moral and ontological approach to virtual-item theft should not forget the concepts of exclusion and value. The charging, jurisdiction and remedy

[16] P.H. Padovan, C.M. Martins & C. Reed, *Black Is the New Orange: How to Determine AI Liability*, 31 ARTIF. INTELL. & L. 69–99 (2023).
[17] A. Bagattini, *Children's Well-Being and Vulnerability*, 13 ETHICS & SOC. WELFARE 211–15 (2019), https://doi.org/10.1080/17496535.2019.1647973.
[18] Litska Strikwerda, *Theft of Virtual Items in Online Multiplayer Computer Games: An Ontological and Moral Analysis*, 14 ETHICS & INFO. TECH. 89–97 (2012).

depend on which of the offenses of non-privileged use for access to computer system, cheating, criminal breach of trust or misappropriation may be applied by prosecutors, if any, on the basis of facts.

A technology neutral, statutory definition of a protected digital asset should acknowledge that a digital asset can be the subject of exclusive control, transfer, valuation or use and that the digital asset is regarded by the parties as having economic or functional value. This does not mean that all game permissions need to be treated as protected digital assets. Courts should take into account the platform agreement, technical architecture, market convertibility and the victim's real loss. Restitution mechanisms should be allowed to return an asset or compensate for an equivalent value, even if criminalization of the confiscation of these sorts of assets is hindered by cross-chain transfers.

4.2 Harassment, Stalking and Immersive Sexual Harm

Persistent following of someone through spaces, threats, disclosure of personal information, the 'recording' of unwanted footage and playing with an avatar's attire, or 'contact with an avatar's sexual parts' may be considered as 'harassing/immersive'. Specific facts may be encompassed by existing offences including stalking, voyeurism, sexual harassment, criminal intimidation, defamation and obscene or sexually explicit publication. But inherited meanings can be based upon physical contact, gender-specific factors, repeated communication, or generation and distribution of content. A single immersive event can occasion grave harm but none of them fit neatly.

There are two extremes in law that should be avoided. Treating any virtual violence as physical violence ignores the statutory requirements and the distinction between simulated and real violence. Treating the event as unreal bypasses embodiment, haptic feedback and psychological harm. A harm and safety concept has been advocated by Chawki, Basu and Choi (2024)[19] which acknowledges the blurred lines between virtual and real world experience. The more effective approach is offence-specific analogy, that is, asking what is the nature of the protected interest that the offence serves: bodily integrity, freedom from fear, protection of reputation, freedom from molestation, sexual autonomy, privacy or other? In each case, the question is whether the metaverse conduct is in some way intended and serious to invade that interest. Legislatures can then fill in tight holes, such as making it a crime to engage in serious, intentional, nonconsensual immersive sex with someone if the perpetrator knows the victim is not consenting, and that the act could reasonably be expected to cause substantial distress in the victim.

Safety by design is required as an accompaniment. Setting personal boundaries to block that immediately removes visibility and audibility, consent prompts to facilitate haptic interaction, default recording notices and accessible reporting can help to prevent or limit harm. These are no excuse for perpetrators or an excuse for the victim to suggest "how could you have been more secure?" Nor should platforms be able to delete evidence in the wake of a report. The interface used for speedy separation should also result in the proportionate preservation of pertinent data.

4.3 Interconnection of Platforms

Platforms exist in an interstice between the roles of venue, identity provider, evidence custodian, market operator and rule maker. In addition to this, intermediary protection is provided by section 79 of the IT

[19] Mohamed Chawki, Subhajit Basu & Kyung-Shick Choi, *Redefining Boundaries in the Metaverse: Navigating the Challenges of Virtual Harm an d User Safety*, 13 LAWS (MDPI) 1–23 (May 2024).

Act[20] if it satisfies the due-diligence conditions and if it does not engage in any prohibited action (such as initiating the transmission, choosing the recipient or altering its contents). Services can have a far more active role in the metaverse than the traditional conduits. They can suggest spaces, create avatars, perform asset transfers, handle haptic interaction and remember spatial data. It is important to determine whether a given service serves as an intermediary or is not an intermediary function by function.

Safe-harbour rules should not protect obstruction, the intentional design of criminal markets or failure to abide by lawful preservation orders, but these rules should not preclude innovation or impose blanket monitoring requirements. A code of practice for the metaverse may include guidelines for risk assessment, child-safety defaults, transparent moderation, emergency reporting systems, retention policies, preservation of data and a designated point of contact for law enforcement. Service sizes, functions and risks should be the basis for graduated duties. Compliance can be tested through independent audits and transparency reports, without the need to inspect all interactions.

5 JURISDICTION IN A BORDERLESS BUT NOT PLACELESS ENVIRONMENT

5.1 Multiple Territorial Connections

A metaverse incident can have at least six locations: the victim, the suspect, the access device, the platform company, the processing server and the wallet or asset service. Not all of them are necessarily in a single country. The concept of territorial jurisdiction does however have its uses as people and buildings are physically located. The issue is not the loss of territory, but concurrent jurisdiction.

The typical methods that states use to assert jurisdiction are based on territorial conduct, effects within the state, the nationality of the offender, the nationality of the victim, and protection of essential state interests. In the case of conduct with a computer, computer system or computer network in India, the IT Act has an express provision of extraterritorial application under section 75. The BNS is also extraterritorial in certain situations and the Bharatiya Nagarik Suraksha Sanhita, 2023 (BNSS) regulates the place of inquiry and trial of offences and also provides machinery for reciprocal assistance. While these rules may create Indian jurisdiction in numerous instances of cross-border events, it is important to note that legal jurisdiction cannot ensure the means to access a foreign suspect or data.

An effects-based approach is appealing in cases of loss or harm suffered by a victim in India, but the unlimited effects jurisdiction has the risk of global over-reach, as virtually any online conduct is deemed to occur everywhere. A substantial connection test should take into account intentional targeting, foreseeable and significant harm, nationality/habitual residence, critical infrastructure location, and ease of evidence. Multiple states may have a legitimate interest, and prosecutors should coordinate based on factors including: gravity of the case, number and vulnerability of victims, suspect location, comparative evidence collection capabilities, fair-trial rights and risk of double jeopardy.

5.2 Terms of Service Are Not Criminal Jurisdiction

Platform terms often specify a forum and a law to govern contractual disputes. The clauses are inoperative as a basis for a state's criminal jurisdiction. Nor, indeed, can a platform claim to be a legally independent virtual district simply by having users opt for the rules of the community. Private ordering can operate as

[20] Section 79 of the Information Technology Act provides a "safe harbour" that protects internet platforms from legal liability for content posted by their users.

a means of regulating acceptable behaviour and the ownership of things; public law remains for crime, coercive investigation and rights.

Problems arise in the identification of the responsible recipient for ordering in decentralization. Although there may not be a central headquarters for a protocol, interface operators, validators, wallet operators, developers and/or governance participants or others in their various capacities may exercise differing forms of control over the protocol. The imposition of liability or disclosure obligations should be based on the actual control and legal ability of a project, not on any mere association with a project. Generalizing to open-source developers would put the brakes on innovation and could be problematic for culpability values.

5.3 Cross-Border Cooperation and Speed

Traditional mutual assistance is usually too slow for volatile and mobile data. The difference between disclosure and preservation: A faster order can "preserve" relevant data, and a slower order can determine whether it may be lawfully "disclosed." This should inform reform. The Budapest Convention's 24/7 Network[21] provides for urgent assistance and preservation of electronic evidence by its participants and its Second Additional Protocol fosters cooperation with service providers and disclosure of electronic evidence. The UN Convention against Cybercrime adopted by the General Assembly in 2024 establishes a more comprehensive basis for cooperation and electronic evidence, but will only be effective once it is signed, ratified, and given effect.[22]

India needs to enhance the existing specialist contact system for availability at all times, and establish a standardized path forward for emergency requests and direct provider contact in line with Indian law. Requests must be specific to the account or session, offence, data categories, time window and necessity. Providers should be cautious to maintain confidentiality and should only disclose accordingly upon Constitutional permission. The dual-criminality requirement can interfere with true innovative harms; states should allow flexibility in how they interpret this based on their underlying conduct and their protected interest, with safeguards against harmful requests or those based on political motives or rights abuses.

Conflicts of law also have an impact on privacy. A disclosure under data protection or constitutional standards of the requesting state may be in contravention of data protection or constitutional standards of the receiving state. Cooperation tools should provide for necessity, proportionality, limitation of purpose, security, deletion and remedies. The users should typically be alerted when there is no longer a need for the seal and providers should be able to contest any order that appears to be unnecessary or too broad.

6 Immersive Digital Evidence

6.1 What Constitutes a Metaverse Crime Scene

A metaverse crime scene is distributed and layered: account registration and login logs, avatar identity and inventory, voice or text communications, 3D scenes, object and collision logs, headset and controller telemetry, gaze and gesture data, device artefacts, moderation, AI prompts and outputs, smart contracts,

[21] Convention on Cybercrime art. 35, Budapest, Nov. 23, 2001, ETS No. 185.

[22] United Nations Convention against Cybercrime, G.A. Res. 79/243 (Dec. 24, 2024), https://www.unodc.org/unodc/en/cybercrime/convention/text/convention-full-text.html.

blockchain transactions, wallet provider logs, and wallet data. It's not the case that one source captures the event in an exact manner.

A victim's recording will include what was recorded on the victim device and not necessarily what was in the system state. Coordinates can be recorded on a platform log without the cues necessary to interpret them. A blockchain will reliably record a transaction, but it will not make some sort of reflection on whether or not consent was obtained fraudulently. Movement of the device could be connected to a controller, but it would not indicate who was operating the controller at the time of the movement. In this regard, Seo, Seok and Lee (2023) suggest a specific metaverse forensic framework, while AlMutawa, Ikuesan and Said (2024) stress the important issues of mapping forensic tasks and technology[23].

Use of a layered reconstruction model by investigators is recommended. The presentation layer contains information on what the user has heard, seen, or felt. The interaction layer is an application of avatars, communications, and spatial events. Platform logic and moderation actions reside in the application layer. The infrastructure layer contains server, cloud and network data. The transaction layer covers blockchains, wallets and payment rails. The physical layer includes devices, local files and the human environment. Multiple layers of corroboration lower reliance on a single proprietary account.

6.2 Volatility and Preservation

Many immersive events are temporary. Voice might not be maintained, an instance might be closed, and an avatar might change, an AI model might be changed or a smart contract might move the assets automatically. Preservation should start as soon as possible but it should be limited in scope. Platforms should include an in-product evidence hold that is triggered through a serious report, a set period of time in front of and behind a report, and metadata of the participants and integrity of the report. The hold should be backed up and subject to expiry and/or legal extension.

There is nothing that can be substituted for forensic readiness, but there is such a thing as "over-retention". Continuous recording of all users' gaze, rooms and intimate movements would create serious surveillance and security risks. Data minimization and evidence preservation are not mutually exclusive, and can be accomplished via short default retention, event triggered holds, local encryption, separation of identity data and strict access control. Preservation should be auditable and criminal investigators should have lawful permission to add intrusive categories.

6.3 Data Pertaining to Authenticity, Integrity and Chain of Custody

Authenticity is the question as to whether it is what its promoter says it is. Integrity: has it changed? Reliability is whether it can be reliably produced. These are overlapping but should not be folded. A correct crypto hash can assure that a file's content hasn't altered since it was collected but doesn't assure that original data was accurately collected. There are no either/or situations with blockchain storage either, as information entered is not validated.

A defensible process should document the collection authority, collector, date and synchronized time source; platform and software version; account, device and session identifier; data schema; how the data was collected; cryptographic hashes; transformations; where the data is stored; who accesses the data

[23] S. Seo, B. Seok & C. Lee, *Digital Forensic Investigation Framework for the Metaverse*, 79 J. SUPERCOMPUTING 9467–85 (2023), https://doi.org/10.1007/s11227-023-05045-1; A. AlMutawa, R.A. Ikuesan & H. Said, *Towards a Comprehensive Metaverse Forensic Framework Based on Technology Task Fit Model*, 16 FUTURE INTERNET 437 (2024), https://doi.org/10.3390/fi16120437.

where; and each transfer. Proprietary formats should be exported in a native format as well as in an open, documented format. Preserved source data for visualization in court should be repeatable with specifications of assumptions and interpolation clearly stated. In addition, Alruwaili (2021)[24] recommends the implementation of distributed chain-of-custody processes; however, technology must complement not supersede the accountable procedure of humans.

Provenance is essential in the case of deepfakes. Experts may have to be called to provide evidence on detection error, compression, artefacts in the model and replay. The party claiming use of the recording is not responsible for proving a negative, but should be responsible for establishing a chain of credibility from source to exhibit. The defence must have adequate access to the underlying data and tools to challenge the prosecution's reconstruction, except in so far as this information is protected by privacy rights orders.

6.4 Evidence and Procedure in India

The Bharatiya Sakshya Adhiniyam, 2023 ("BSA")[25] expressly includes electronic and digital records as documentary evidence and lays down conditions for admission of electronic records. It could include records on the platform, wallet transactions, data from devices, and immersive recordings. A critical legal question will typically be whether the electronic record is in compliance with the statutory requirements that specify the way in which an electronic record is to be proven, and the reliability of the system that generated it. A screenshot or exported video can never be considered to be the legally relevant original by investigators.

The BNSS modernizes a number of aspects of the procedure, such as the acceptance of electronic communication and audio-video in criminal process. Search, seizure, forensic examination and reciprocal assistance powers can be used to aid investigation of metaverse cases. However, a traditional mindset about devices and seizures will not suffice if a key record is stored in the cloud somewhere other than your jurisdiction or on an out-of-control distributed ledger. Operational protocols should clearly define the process used to request emergency hold, capture virtual scene live, image off a headset, get a platform schema, trace a wallet, and document software dependencies.

Specialised forensic labs and trained examiners are required. Teams might need specialists in the area of virtual reality, network forensics, blockchain analysis, AI provenance, trauma-informed interviewing, and comparative law. The problem of interpretative error is also prevented by a multi-disciplinary approach. An avatar animation could be one that the platform generates automatically—or one the user selects; the forensician can describe the animation system, and investigators can evaluate communications around that intent.

6.5 Privacy, Privilege and Due Process

Metaverse evidence can reveal bystanders, private residences, health issues, and intimate behaviour and relationships. Religious items or medication could be found during a room scan taken as part of a play boundary. The gaze data can reveal interests or impairment. A broad warrant for an entire history of an

[24] F.F. Alruwaili, *CustodyBlock: A Distributed Chain of Custody Evidence Framework*, 12 INFORMATION 88 (2021), https://doi.org/10.3390/info12020088.

[25] Sections 57 and 61 classify electronic and digital records as valid documents and primary/admissible evidence on par with physical records.

account is thus disproportionate. Orders must be subject to time limits, the number of participants, data type and alleged offence with special justification for biometric and spatial-home data.

Privileged communications and irrelevant third parties should be protected in filtering. Investigators should capture results of automated analytics and report on material error rates. Any indication of suspicious gesture(s), match of voice(s) or reconstruction of a scene by AI must be treated as an investigative lead unless independently validated. The accused should be able to adequately challenge the model, the data and any inferences made[26]. There are also issues of equality as the biometric and speech systems could function differently between populations.

Victims need control and support. Trauma-informed collection should involve as little re-exposure to recordings as possible, provide clear options regarding sensitive content, and clarify disclosure in court proceedings. At the same time, the right to safety cannot be deemed in and of itself a proof of guilt, because of the right to due process. Criminal adjudication is based on separate evidence that is tested substantially in accord with law.

7 A COORDINATED REGULATORY MODEL

The analysis suggests a multi-layered approach in the form of statutes instead of a single metaverse statute. It consists of five layers, the first being substantive criminal law. Offences already defined as technology neutral should be applied as appropriate if the elements of these offences really apply. Legislatures should only include an exception for specific problems or narrow exceptions that have been proven, specifically, and especially, digital-identity interference and protected digital assets and intentional non-consensual immersive sexual conduct. Drafting should be directed towards control, human influence, deception, consent, harm and culpability, and not at particular brands or equipment.

The second layer is procedural readiness. The law should differentiate between urgent preservation and disclosure, provide for limited and targeted cross-border requests, and set up expedited judicial review. A 'standard metaverse evidence-preservation notice' stating identifiers, relevant time periods, categories and retention time period could be implemented in India. There should be a designated 24/7 contact point for coordination with platforms and foreign counterparts, and have an open log of requests and results.

Forensic standardization is the third layer. Minimum acquisition and chain of custody fields, for immersive evidence, should be defined in a national protocol, including time synchronization, coordinate systems, software versions, rendering parameters, AI involvement and cryptographic integrity. Export formats should be interoperable and include data dictionaries. VR scene reconstruction and blockchain analysis should also be assessed and accredited, in addition to proficiency testing. The protocol should be created in consultation with forensic institutes, courts, law enforcement and defence lawyers, platforms and civil-society experts.

The platform governance layer is the fourth layer. Services should be designed to be safe, have multiple tiers of identity assurance, evidence holds, accessible reporting and child protection, along with a lawful-request interface. There is a link between working practices and risk. It isn't fair to impose

[26] P. Gulati & K. Pal, *The Notion of Privacy Rights in the Metaverse: Examining Legal Hemispheres in India's Digital Era*, *in* METAVERSE DRIVEN INTELLIGENT INFORMATION SYSTEMS (S. Kautish & Á. Rocha eds., 2024) (Information Systems Engineering and Management, vol. 20), https://doi.org/10.1007/978-3-031-72418-3_10.

compliance obligations on a small creative world that would apply to a big financial marketplace or kids' space. Safe harbour should remain for good-faith intermediaries, but responsibilities for transparency and preservation should come into play for functions where a platform has control over risk and evidence.

The fifth layer is rights oversight. Intrusive identification and access to data should be supervised by independent authorities and courts. There should be an aggregate transparency report that includes the number of preservation and production requirements, number of emergency requests, rejection rates and transfers across national borders. There should always be remedies for wrongful disclosure, retention of identity over and above the need and retention beyond lawful need. To mitigate the false choice between accountability and anonymity, available privacy-enhancing technologies include selective-disclosure credentials, encrypted identity escrow and zero-knowledge age proofs.

In the international context, elements of harmonisation should focus on the procedure and on the minimum offence concepts and not on uniform cultural regulation of all virtual behaviour. Agreements can be made regarding expedited preservation, points of contact, authenticity documentation, asset freezing, and asset safeguards despite any differences in substantive laws. The Budapest system provides lessons for operation and the UN Convention against Cybercrime provides a potentially broader ground for cooperation. There should be a review of the implementation of these in light of human rights, since the scope of cybercrime powers could be used for surveillance or suppression of expression.

Lastly, regulation should be an ongoing process. Mandatory incident reporting and anonymised empirical studies can tell us what harms do, and how harms are not addressed through platform controls, or by prosecution. Speculative offences can be prevented from becoming permanent via sunset clauses or periodic review. Regulatory sandboxes can be used to trial age-assurance mechanisms, evidence export and privacy-preserving identity systems — but only under regulatory guidance. The objective is to be adaptable at all institutions and not to predict all future virtual worlds.

8 Conclusion

Embodiment, persistent identity, virtual economies and pervasive sensing exacerbate vulnerabilities already seen in digital criminal justice systems in general in the metaverse. The key legal issues are who is responsible, jurisdiction, and obtaining adequate evidence. Avatars can be a useful tool for helping to bring people to life, but they will not create a new legal person. A virtual asset need not be regarded as property in order to be protected. A violation, even if it is not literally translated as physical contact, can be very serious and cause significant damage. The differences can help keep law principled amidst technological change.

The IT Act, BNS, BNSS and BSA offer a good base for India. They can respond to unauthorized access, identity theft, personation, cheating, stalking, intimidation or illegal content/digital records. However, enforcement efforts will continue to be uneven, with the absence of clarity in the law and practice on digital identity interference, virtual-asset control, immersive sexual harm, platform preservation and cross-border access. Reforms that are functional and procedural – graded identity assurance, a significant connection regime, prompt preservation that is subject to review, standardized multimodal forensics, and duties on the platform based on risk – are the most useful.

Accountability is possible without the omnipresence of identification. The values of lawful pseudonymity, experimentation, and intimate association deserve to be protected, and the metaverse can enable

these values. A legitimate structure will link serious wrongdoing with responsible individuals using a proportionate process, rather than through the blanket of surveillance. Technologically neutral criminal law, focused cooperation of specialists, forensic precision and protection of rights can help shield victims, while respecting the rule of law and due process, as 3D environments become a commonplace part of social and economic interaction.

* * *

PART II

Intellectual Property and Generative AI

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla
(VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.9

CHAPTER 9

Intellectual Property in the Age of Generative Technology: Rethinking Authorship, Training Data, and Infringement Liability

DR. SHYAMAL M. DAVE[*] AND ADITI PATEL[**]

Generative AI systems are trained on very large bodies of copyrighted text, images and code, and whichever AI system is used, e.g. ChatGPT or DALL-E, raises another copyright question: who, if anyone, owns what the system produces? This paper investigates both extremes of this issue. As for inputs, it compares the methods of ingesting copyrighted works in training generative models: how the United States courts are starting to split in their decision and how, without a text-and-data-mining exception, Indian law remains significantly less clear than, for example, the EU's opt-out model under the AI Act. On the output side, it considers the requirement of human-authorship, now affirmed at the highest level of the U.S. courts to date, and whether this requirement is inescapably dictated, or simply permitted, by an originality standard under the Indian copyright law. The core of the paper is that India is at a genuine tipping point: It was not bound by any precedents on either question, there was an ongoing case before the Delhi High Court, and there was a government paper proposing a compulsory blanket-licensing system that had not yet been passed. The paper shares some lessons from the recent developments in the case law of the United States, Europe's regulatory framework, and India's shifting policy stance, and suggests that a hybrid model between full (or semi-) statutory licensing and reliance on fair-dealing litigation is more desirable than leaving the issue for the courts. It ends with concrete suggestions on how the Copyright Act, 1957 can be changed to cover training-data licensing and the question of the authorship of AI-assisted works.

Keywords: Generative Artificial Intelligence; Copyright; Text and Data Mining; Authorship; Fair Use

1 INTRODUCTION

The field of generative artificial intelligence (AI) has transitioned from the realm of research to infrastructure over the last few years. Prose, software, images, music, you name it, is now written on command by large language models and diffusion models, because they were first trained on very large amounts of existing human-generated text – much of which was copyrighted, and almost none of which explicitly licensed the use of the model's output at the time the original model was created.

This paper focuses on the two copyright issues that generative technology has opened up. The first is an "input question": Is copying a copyrighted work into a training dataset, and using that to adjust the model's parameters, a violation of the reproduction right, or is it otherwise excused, for example by fair use/fair dealing, etc.? The second is an output question: When using a generative system to create an image, a poem, or code with little or no human creative involvement other than as a prompt, can anything at all be copyrighted, much less by the generator?

[*] Assistant Professor at Department of Law, Parul Institute of Law, Parul University, Vadodara, Gujarat, India.
[**] Assistant Professor at Department of Law, Parul Institute of Law, Parul University, Vadodara, Gujarat, India.

Both questions have been the subject of ongoing litigation and regulatory activity in key jurisdictions, and responses are heading in opposite directions with important implications for Indian law and Indian creative industries. This paper examines the state of the law in the USA, where the first substantive judicial decisions on both topics have been issued; in the European Union, where a different regulatory model (opt-out) exists within the framework of the EU AI Act; and in India, where the issue continues to be doctrinally unresolved and where a government policy proposal is pending, as is litigation before the Delhi High Court. The paper asserts that the sticking to the undefined fair-dealing principle by India is a "shaky long-term strategy" and that the better strategy would be to adopt a "calibrated statutory licensing model."

2 RESEARCH PROBLEM AND RESEARCH METHODOLOGY

The problem of research that is tackled here is two-fold. First, the current copyright doctrine was developed with a different model of authorship and discrete copying in mind; it is not easily adaptable to technology that processes millions of works at once and outputs seemingly independent but nonetheless similar works. In the case of India, there is no text-and-data-mining exception in the copyright law that exists in the European Union and no final judicial ruling as to the scope of Section 52 fair dealing to AI training. Secondly, the Copyright Act, 1957, which governs copyright protection for computer-generated works, provides a definition of an 'author' to be the 'person who causes the work to be created'. However, this fails to distinguish whether a certain amount of human involvement is necessary or not to constitute the work as being made by a 'person', especially where a generative system performs most of the expressive act.

This is a comparative and doctrinal paper. It explores key legal texts, such as the Copyright Act, 1957; the United States Copyright Act; Regulation (EU) 2024/1689 (the EU AI Act); and Directive (EU) 2019/790, and evaluates recent case law from the United States and existing Indian case law and policy developments in light of this comparative framework. It does not conduct an empirical analysis of licensing markets or of the designs of licensing architectures; instead, its findings are of a legal-doctrinal nature.

3 THE INPUT PROBLEM – TRAINING DATA AND THE RIGHT OF REPRODUCTION

3.1 Looking for a Place to Start?

Typically, training a generative model involves copying the training corpus, at least temporarily, and putting it in a model's format, and this requires implication of the reproduction right under most copyright laws, such as Section 14 of the Copyright Act, 1957. The most prevalent defence advanced by AI developers is that this copying is purposeful; that the AI model is neither creating copies for storage nor an attempt to retrieve the original work, but rather is learning some statistical patterns of the original, and that the resulting system is put to a use, such as generating new content, that is not the same as the use of the original as, for example, a news article or photograph.

3.2 The United States: A Doctrine Starting to Crack

In Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., the District of Delaware granted a renewed motion for summary judgment, finding that Ross Intelligence's use of Thomson Reuters'

copyrighted headnotes to train a rival legal-research tool was not a fair use.[1] Significantly, the court noted that Ross's use was commercial, that it did not involve any meaningful alteration of the material copied, and the significance of market substitution, meaning that such a use was designed to take advantage of the derivative market, for the purpose of teaching competing databases how to function and that Ross's product was intended to compete directly with Westlaw.[2] While commentators have been careful to point out that Ross's tool was not itself generative and may not be broadly applicable to large language models generating novel content as opposed to excerpts, the ruling is the first judicial setback for the transformative-use argument in the AI-training landscape and is likely to be referred to frequently with respect to the pending generative-AI cases.[3]

The pending case of greatest interest is perhaps The New York Times Co. v. Microsoft Corp. in which the Times claims that millions of its articles were used without permission by OpenAI and Microsoft to train GPT models that generate content that rivals its own journalism.[4] In April 2025, the district court denied the defendants' motion to dismiss on most of its terms, leaving the Times' central claims to proceed to discovery.[5] In turn, the case was joined with parallel publisher suits in multidistrict litigation and, at this time, no trial date has been set; the resolution of this case is expected to set the tone for the entire generative-AI industry's exposure to claims involving training data.[6]

There is another parallel (sort of) line of suits involving models that produce images. The plaintiffs in Andersen v. Stability AI allege that Stability AI, Midjourney and DeviantArt used copyrighted images to train image models without permission, and Getty Images is suing Stability AI in Delaware for the use of millions of its licensed images, associated captions and metadata, to create Stable Diffusion.[7] Both cases are still in discovery – and the fact that only the thinnest of margins was found in Getty's favor in the parallel UK case brought by Getty against Stability AI, highlights how unclear the doctrine is even within one litigant's parallel litigation proceedings across jurisdictions.[8]

3.3 The European Union: Opt-Out Approach to Regulation

The European Union has attempted to deal with the same issue in the legislature rather than on an individual basis. The concept of a text and data mining exception, which already exists in Article 4 of Directive (EU) 2019/790, applies only where rightsholders have not expressly reserved their works in an appropriate manner, such as by machine-readable means.[9] The EU AI Act (Regulation (EU) 2024/1689) adds that providers of general-purpose AI models must implement a policy, thus complying with the EU copyright law, to respect any opt-out reservations and provide publicly a sufficiently detailed summary of

---

[1] Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., No. 1:20-cv-613-SB, 2025 WL 458520 (D. Del. Feb. 11, 2025).
[2] *Id.*; *see also Court Rejects Fair Use Defense in AI Copyright Case*, GOODWIN (Feb. 2025), https://www.goodwinlaw.com/en/insights/publications/2025/02/alerts-practices-ip-lit-court-rejects-fair-use-defense-in-ai-copyright-case.
[3] *Training on Thin Ice:* Thomson Reuters v. ROSS *and the Future of Fair Use for AI Systems*, CTR. FOR ART L. (Oct. 2025), https://itsartlaw.org/art-law/training-on-thin-ice-thomson-reuters-v-ross-and-the-future-of-fair-use-for-ai-systems/.
[4] N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023), *consolidated as In re* OpenAI, Inc., Copyright Infringement Litig., No. 1:25-md-03143 (S.D.N.Y.).
[5] NYT v. OpenAI *Lawsuit Status 2026*, AI VORTEX (May 19, 2026), https://www.aivortex.io/legal/ai-case-law/nyt-v-openai/.
[6] *AI Copyright Training Data Lawsuits 2026: Status, Timeline, Risk*, AI VORTEX (May 19, 2026), https://www.aivortex.io/legal/guides/ai-copyright-training-data-2026-landscape/.
[7] Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal.); Getty Images (US), Inc. v. Stability AI, Inc., No. 1:23-cv-00135 (D. Del.).
[8] *The Copyright Battle Over AI-Generated Images: Where Things Stand in 2026*, ANALYTICS INSIGHT (2026), https://www.analyticsinsight.net/news/the-copyright-battle-over-ai-generated-images-where-things-stand-in-2026.
[9] Directive 2019/790, art. 4, 2019 O.J. (L 130) 92 (EU).

content used to train the model.[10] It is a very different structure than the "after the fact" fair use litigation in the U.S.: it assumes training is legitimate but allows rightsholders to exercise an affirmative right to withdraw their works from the use of training provided that it is accompanied by a transparency document so that the rightsholders can verify it.

3.4 India: An Unresolved Question

There is no specific text-and-data-mining exception in Indian copyright law. The Copyright Act, 1957 provides an exhaustive list of the various instances of fair-dealing in Section 52, and it has yet to be categorically determined whether AI training of a commercial model would fall under any of them, such as under the protection of private use, research, or incidental use.[11] That is exactly the Delhi High Court's question in ANI Media (P) Ltd. v. OpenAI, Inc., and the news agency ANI has declared that it claims that OpenAI accessed its copyrighted news content without a licence, to train ChatGPT, and that OpenAI itself, similar to its US counterparts, has registered that no harm occurs in India, since the servers on which the training took place are located abroad. The final judgment will be the first definitive judicial pronouncement on the applicability of Indian fair dealing to AI training; and, because of the cross-border reach of Indian copyright law contested by OpenAI, will give important guidance on the cross-border reach of Indian copyright law in the AI context.[12]

While the courts are still deliberating, the executive has taken a separate policy route. In December 2025 the Department for Promotion of Industry and Internal Trade published a working paper, which has proposed a 'hybrid' licensing model, where the AI developers may be granted a "licence of sorts" to use copyright content to train AI and a centralized royalty collection body (provisionally named the Copyright Royalties Collective for AI Training) to pay royalties to rightsholders.[13] Notably, the working paper's opt-in 'blanket-licence' approach (as opposed to the opt-out right in the EU's text-and-data-mining exception provision) for rightsholders was not widely accepted: although industry bodies separately dissented from certain parts of the proposal on cost basis, creator groups criticised the working paper's opt-in approach in its current form.[14] The working paper, at this time, is merely a policy proposal because as yet, no Bill to amend the Copyright Act, 1957 has been tabled on the Parliamentary floor and the Ministry of Commerce and Industry has also set up an eight-member expert committee to review the adequacy of the Act in view of generative AI technology.[15]

---

[10] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence, art. 53(1)(c)–(d), 2024 O.J. (L, 1689).

[11] Copyright Act, 1957, § 52.

[12] ANI Media (P) Ltd. v. OpenAI, Inc., CS(COMM) No. 1028 of 2024 (Del. H.C.); *see The Future of Indian Copyright Legislation in the Wake of* ANI Media v OpenAI, WORLD TRADEMARK REV. (2026), https://www.worldtrademarkreview.com/guide/india-managing-the-ip-lifecycle/2026/article/the-future-of-indian-copyright-legislation-in-the-wake-of-ani-media-v-openai.

[13] DEP'T FOR PROMOTION OF INDUS. & INTERNAL TRADE, WORKING PAPER ON ARTIFICIAL INTELLIGENCE AND COPYRIGHT (PART I) (Dec. 2025); *AI Training & Copyright in India: DPIIT Hybrid Model Explained*, INTEPAT (May 13, 2026), https://www.intepat.com/blog/ai-training-copyright-india-dpiit-hybrid-model.

[14] *The Flaws in India's Approach to Copyright Reform for AI*, LIVE LAW (Jan. 31, 2026), https://www.livelaw.in/articles/copyright-reform-ai-521269.

[15] *The Future of Indian Copyright Legislation in the Wake of* ANI Media v OpenAI, *supra* note 12.

4 AI-GENERATED WORKS: THE OUTPUT PROBLEM

A more clear judicial response has been given (at least in the United States) to the second question: Can the "creations" of a generative system be the subject of copyright protection, and if so, whose?[16] In March 2026, the Supreme Court declined to grant certiorari in Thaler v. Perlmutter, and the human-authorship requirement is now firmly established in the D.C. Circuit at the appellate level — at least on the specific facts of work claimed to have no human creative contribution at all.[17]

The ruling doesn't answer the more difficult and frequent question: is there sufficient human intervention – in prompting, selection, editing – in the creation of an AI's answer to justify a claim that, at least for those parts, humans acted as the author? The United States Copyright Office's own guidance has suggested a distinction between merely prompting a system, which is not sufficiently creative to establish a claim to authorship, and material that relies on a human author's more significant selection, arrangement, or modification of the AI-generated material, in which case, authorship for those parts of the work may be asserted.[18] That line-drawing effort—whether assistive use or just prompting—will likely supplant registration practice and litigation for the foreseeable future, and a related case, Allen v. Perlmutter, is likely to test whether the content is assistive use or just prompting in a context more similar to typical prompt-based generation than to the fully autonomous creation claimed by Dr. Thaler.[19]

This issue has yet to be put to judicial test in India. The Copyright Act, 1957 already provides Section 2(d)(vi) for computer-generated works, which is peculiar to jurisdictions including India as it defines the author of a literary, dramatic, musical or artistic work made by computer as 'the person who causes the work to be created'.[20] This provision was written with computer-assisted—not computer-autonomous—creation in mind, however, and the words of the statute do not imply copyright functionality for the causing person's contribution, just causative. No reported Indian decision has tested whether the necessary attributes to constitute a generative-AI prompt satisfy this standard, or even whether the provision has the capacity to support the complete autonomy of large-language-model output at all, and the paper contends that this is equally one of the issues that needs to be addressed either by judicial determination or legislative amendment which provides a minimum standard of human creative participation.

5 EVALUATING REGULATORY MODELS: LITIGATION, EXCEPTION, OR LICENCE

The jurisdictions surveyed are now known to have three general regulatory models. The first, illustrated in the United States, allows decisions to be made case by case in the courts, resulting in fact-sensitive, slow, and – as the contrast between the Ross Intelligence decision and the still-pending New York Times and Getty Images accounts demonstrates – uncertain convergence even in the same legal system. The second type, as with the European Union, sets up a default permission with a far lower hurdle to conform (called opt-out) but also a lesser burden for rightsholders to opt out and a transparency requirement on the AI developer. The third one (proposed, but not yet in effect in India) makes an 'opt-out' and individual litigations redundant and provides a compulsory blanket licence based on a centralised collecting

[16] Thaler v. Perlmutter, 130 F.4th 1039 (D.C. Cir. 2025).
[17] Thaler v. Perlmutter, 130 F.4th 1039 (D.C. Cir. 2025), *cert. denied*, No. 25-449 (U.S. Mar. 2, 2026).
[18] U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 2: COPYRIGHTABILITY (Jan. 2025).
[19] Thaler v. Perlmutter*: Human Authors at the Center of Copyright?*, KLUWER COPYRIGHT BLOG (Apr. 8, 2025), https://legalblogs.wolterskluwer.com/copyright-blog/thaler-v-perlmutter-human-authors-at-the-center-of-copyright/.
[20] Copyright Act, 1957, § 2(d)(vi).

society as is provided in the statutory licensing schemes already in existence under Section 31D of the Copyright Act, 1957 for broadcasting.[21]

Every model has varying risk distributions. Litigation-based fair use is unstable for those looking to develop and slow for those whose work is used without their permission to receive payment, but retains the right of the individual rightsholder to outright refuse a licence in the proper case. The opt-out model recently adopted by the European Union helps to limit litigation risk for a developer who has applied the opt-out reservation and transparency requirements, but it may also shift responsibility on to the individual rightsholder that may not have the capacity or awareness to activate the opt-out – the majority of whom are outside of the large publishing houses. This paper considers the blanket-licence model without an opt-out—which is proposed in India to be the system of choice for the country—to be the system's main weakness as it prioritises AI development without even considering a rightsholder-level opt-out measure.

6 RECOMMENDATIONS

The DPIIT's hybrid licensing proposal must first be improved before it's passed by having a rightsholder opt-out clause, similar to the EU's Digital Single Market Directive, to make the statutory licence, instead, a default rule. But without an opt-out provision, a blanket licence will be vulnerable to constitutional scrutiny as an unreasonable interference with the right to property in copyright (under Article 300A of the Constitution) and it would also prove irreconcilable with the obligations of India in the Berne Convention's three-step test requiring exceptions to be confined to "certain special cases" which do not "conflict with a normal exploitation of the work."[22]

Second, Section 2(d)(vi) of the Copyright Act, 1957 could be amended by adding a clear definition of what is a 'computer-generated work' and by clearly specifying that it can only be a 'computer-generated work' where the human contribution is more than a 'natural-language prompt' in line with the 'assistive use' and 'mere-prompting' distinction evolved by the United States Copyright Office, to prevent courts from formulating a definition of the phrase from first principles when the question will certainly arise before them.

Third, require the proposed Copyright Royalties Collective for AI Training to conduct transparency reports in the form of Article 53 of the EU AI Act to ensure that Indian creators can at least have visibility at the 'dataset' level if their works have been used by the Collective or not and/or how they have been used, rather than just the 'hot-potato' of the royalty payments where the calculation methodology is not transparent and disclosed in the current working paper.

Fourth, as the ANI Media litigation may yet decide the fair-dealing issue before any legislative change is made, any pending decision by the Delhi High Court on this issue should be closely watched as it addresses the territoriality argument – if truly, there is a decision that training of AI models on foreign servers is not subject to Indian copyright jurisdiction, then this ruling, if adopted, would virtually render any copyright regime on such training ineffectual for AI developers with the ability to organise their training on foreign platforms.

---

[21] *Id.* § 31D.

[22] Berne Convention for the Protection of Literary and Artistic Works art. 9(2), Sept. 9, 1886, as revised at Paris, July 24, 1971.

7 CONCLUSION

Generative AI has brought attention to two long maligned copyright holes—the first one pertaining to the limits of what is called a "fair use" to create something new and healthier—a new kind of tool—and the second one has to do with the authorship of whatever that tool creates. Those questions are being answered in court in the United States, and there has been an early decision — Ross Intelligence's fair-use defence was rejected, and Thaler's human-authorship requirement was upheld — that the developers of AI might not have hoped would be the case. The EU has opted for a regulatory, opt-out type of response, which gives way for some administrative burden in exchange for regulatory certainty. In India there is a pending case in the Delhi High Court[23] in which only an interim, prima facie ruling has been given so far, a working paper that impacts works generated by a computer machine, yet does not include an opt-out clause for a supporting work, and a 1957 statute, amended in 1994 with computers but not LLMs in mind. The authors of this paper have argued that India should not wait for the courts to weigh in on how to interpret and apply this technology; it cannot be simply worked around with the current interpretation of Section 52 of the Copyright Act, 1957; nor should it merely adopt DPIIT's proposal for blanket licences.

* * *

[23] ANI Media (P) Ltd. v. OpenAI OPCO LLC, 2026 SCC OnLine Del 5291 (Del. H.C. July 24, 2026) (refusing an interim injunction and holding, prima facie, that the Court had jurisdiction although training took place on foreign servers and that storing ANI's works to train ChatGPT fell within Section 52(1)(a)(i); the observations are expressly without prejudice to the final adjudication of the suit).

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.10

CHAPTER 10

A Comparative Legal Analysis of the Emerging Right to Know Concerning the Use of Copyrighted Works in AI Training Data

C. SOPHIA JEYAKAR*

The rapid development of generative artificial intelligence (AI) has intensified concerns regarding the use of copyrighted works in training datasets, particularly where authors and right holders have limited or no knowledge of whether their works have been accessed, copied, processed, or incorporated into AI training systems. This raises an emerging legal question: whether authors and copyright holders should possess a legally enforceable right to know whether, and to what extent, their copyrighted works have been used as AI training data, and whether existing copyright law provides adequate mechanisms to secure such transparency. The research examines the tension between the commercial and technological interests of AI developers and the informational interests of copyright holders, with particular focus on dataset transparency, disclosure of training-data sources, and the evidentiary difficulties involved in establishing the use of particular copyrighted works.

The study adopts a comparative doctrinal legal methodology, analysing statutory provisions, judicial decisions, regulatory developments, policy documents, and scholarly literature concerning copyright and AI training data. It comparatively examines the emerging approaches in India, the European Union, the United States, and selected other jurisdictions, with particular attention to transparency obligations and copyright-based claims concerning AI training. The paper argues that although a general statutory "right to know" concerning AI training datasets is not yet uniformly recognised, existing copyright principles, transparency obligations, and emerging regulatory approaches provide a foundation for developing such a right. It proposes that meaningful dataset transparency should form an essential component of the contemporary copyright framework by enabling authors to identify potential unauthorised uses, exercise available rights, seek appropriate remedies, and participate more effectively in licensing and remuneration mechanisms. The study therefore advocates a balanced transparency framework that protects copyright interests without imposing disproportionate disclosure obligations that could undermine legitimate AI innovation, trade secrets, or technological development.

Keywords: Artificial Intelligence; Copyright Law; Right to Know; AI Training Data; Dataset Transparency

1 INTRODUCTION

1.1 Background: Generative AI and Copyrighted Training Data

The quick advancement of generative artificial intelligence (AI) has changed the way in which information, creative works and other kinds of digital content are gathered, processed and reproduced. Today's generative AI systems, especially large foundation models, are created through training procedures that make use of large datasets consisting of text, images, audio, video, software code and other types of information. The size of these datasets has fundamentally affected the connection between copyright law and technological progress. Unlike traditional forms of digital use, the development of AI can involve

* Assistant Professor & Research Scholar at Vels Institute of Science, Technology and Advanced Studies, Chennai, Tamil Nadu, India.

the systematic handling of huge amounts of works, many of which come from publicly available websites, digital libraries, repositories, online platforms and other web sources. AI training datasets can include works that are the subject of copyright, such as books and academic publications, newspaper articles and material from the web, photographs and artistic works, musical and audiovisual material, computer programs and databases. Yet the fact that a copyrighted work is included in a dataset does not on its own mean that copyright infringement has taken place. The issue of legal importance is what actions are carried out in obtaining, processing and using the work and whether or not those actions fall within the exclusive rights of the copyright owner or an applicable statutory exception or limitation.[1]

It is especially important to make this distinction since the technological process of training an AI involves a number of possibly different stages. Getting access to a work simply means obtaining or looking at the information it contains, while downloading or copying might involve reproducing the work or a substantial part of it. The act of processing or extracting data from a work may include computational analysis, transformation or extraction of the information, depending on the type of work and the technology used. Training an AI model then constitutes an additional stage in which the information that has been extracted or processed is used to establish statistical or computational relationships inside the model. Lastly, output generation takes place when the trained model produces new content in response to a user's prompt.[2] The various stages should not be regarded as legally equivalent just because they are part of the same technological process. The fact is important since copyright law usually controls specific acts in connection with protected works rather than merely the presence of information in a technological system. When it comes to the legal interpretation of AI training, it is necessary to consider whether a protected work has been reproduced, communicated, adapted, extracted or in some other way used in a manner which falls within the copyright owner's exclusive rights. At the same time, the analysis has to take into account statutory exceptions, limitations, permitted uses and the new approaches to computational analysis and text and data mining.

Since copyrighted material is being used more and more in the development of AI, a complicated balance arises between technological innovation, the need for copyright protection and the legal right to access data.[3] The purpose of copyright is to give authors real control and economic advantages with respect to their creative works, while the development of AI may require access to large and varied amounts of information. It is possible that overly strict measures could hinder technological research and innovation, and on the other hand, there is a risk that the commercial use of copyrighted works without restriction could damage the economic and informational interests of authors and rights holders. In this context, the key question is not merely whether copyrighted works are included in the AI training datasets, but whether authors and rights holders should be able to find out that their works have been used, identify the way in which they have been used and be able to exercise meaningful legal rights in relation to such use. This distinction forms the basis for looking at the new right to know about copyrighted works that have been incorporated into the AI training process and the various approaches taken by India, the European Union and other regions.

---

[1] U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING (2025).

[2] Regulation 2024/1689, of the European Parliament and of the Council of June 13, 2024, Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), recital 107, 2024 O.J. (L 1689) 1 (EU).

[3] Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 743–44 (2021).

1.2 The Problem of Information Asymmetry and of the Evidence

The incorporation of copyrighted materials into the training of AI systems presents a clear issue in the form of information asymmetry between copyright holders and the developers of the AI. While copyright owners generally have knowledge of the works themselves, they often have very little or no reliable information concerning what happens to those works after they have been made available online or otherwise come within the reach of data-collecting systems. It is therefore possible that an author will know that a certain book, article, photograph, work of art, musical piece or software program exists but will not know whether it has been collected, included in a training dataset, processed as part of the AI development process, or finally used in training a specific model.

This lack of information can occur at a number of different stages. A copyright holder might not know whether a work has simply been accessed, whether a copy of it was made during the process of data collection, which dataset or data repository the work was in, whether the work survived the various filtering and preprocessing steps, or whether it was in fact used in training. Even if a copyright owner suspects that a work has been used, it may be hard to identify the exact AI model that was trained on it. The difficulty is increased by the huge size and constantly changing nature of current training datasets, since these may include millions or billions of individual data items gathered from a number of different sources and then subjected to a series of collection, filtering, deduplication, classification and transformation processes. On the other hand, AI developers and the organisations taking part in model development may have considerably more information about the origins and the way in which the training data was handled. According to their internal practices, they might keep dataset documentation, records of data acquisition, logs relating to filtering and preprocessing, provenance information, licensing records and other technical documentation connected with model development. This kind of information could allow a developer to find out if a particular work got into a dataset and how it was then treated. Nevertheless, the existence, availability and evidentiary value of such records can differ greatly from one developer to another.[4]

The asymmetry produced presents a major problem when it comes to evidence. A copyright holder who wants to enforce their rights cannot readily prove a fact which is probably in the knowledge or control of the AI developer, namely, whether their work was actually gathered and used during the training process. Simply because an AI-generated work is similar to an existing copyrighted work does not mean that the work was included in the training data. Likewise, the fact that a work is available to the public does not necessarily show that it was downloaded, copied, processed or incorporated into a specific training dataset. In order to verify this independently, access to technical details might be needed, details which the copyright holder does not have. The implications of this problem go beyond legal proceedings. Since they do not have reliable information about how their works have been used, copyright owners may find it difficult to decide whether or not to take action, to negotiate licences, to seek payment or to challenge certain types of AI development. Information about actual use could also be important for collective licensing and for enabling authors and rights holders to make well-informed commercial decisions about their works. Therefore, the lack of transparency can have an effect not only on the enforcement of current rights but also on the ability to exercise those rights in a meaningful way.

[4] Regulation 2024/1689, recital 107, 2024 O.J. (L 1689) 1 (EU).

The way in which evidence is handled therefore changes the current discussion about copyright and AI training. It is not just a matter of determining whether the use of copyrighted works for training amounts to infringement under the relevant law; an earlier and possibly decisive issue is how a copyright owner can prove that such use has taken place. Since important information regarding data acquisition, provenance and model development is held by the AI developers, the question of who should bear the burden of providing this information becomes a significant legal issue.

The debate now developing around a right to know should not be seen simply as a call for general corporate transparency, but rather as a possible means of dealing with the particular structural imbalance between the people who create copyrighted works and those who build AI systems by using large-scale datasets. The main argument of this paper is that the copyright issue is not just a matter of whether training AI constitutes infringement; it also involves determining who is responsible for proving that copyrighted works were indeed used. This difficulty in providing evidence forms the basis for investigating whether the current copyright systems offer adequate protection for authors and rights holders and whether other legal systems are tending towards imposing specific transparency or disclosure requirements regarding the training data used by AI systems.[5]

1.3 The "Right to Know" Concept

The growing idea of a "right to know" is the possible legal claim of a copyright holder to get enough information in order to assess whether or not their copyrighted work has been used in the training of AI systems, and where appropriate, how it has been used. This notion has become more important since copyright owners often have no knowledge of whether their works have been gathered, included in training datasets, or employed in the development of specific AI models. In the absence of such information, it becomes difficult to put their copyright rights into practice.

It is necessary to differentiate the right to know from other similar concepts. The right to access information relates to the possibility of getting hold of the relevant records or data, while the right to know refers to the fundamental interest in ascertaining whether a work was used. Transparency, on the other hand, broadly speaking, involves obligations on AI developers to make available information about their systems, datasets or training methods. Yet general transparency does not allow an individual copyright holder to check that a particular work was used. Attribution means giving recognition or credit to the creator, and remuneration refers to the payment of financial compensation in respect of the use of a copyrighted work.[6] Thus, the right to know should not automatically be understood as a right to payment, attribution, consent, or a finding of infringement. Rather, it represents an emerging informational entitlement concerning the use of copyrighted works in AI training, which may enable copyright owners to assess and exercise their existing legal interests.

1.4 Research Questions

a) Whether existing copyright laws provide copyright holders with an enforceable right to know whether their works have been used in AI training?

---

[5] *Id.* art. 53(1)(d).
[6] *Id.*

b) Whether emerging transparency obligations in different jurisdictions effectively enable copyright holders to identify such use?

c) Whether disclosure requirements should extend to individual copyrighted works or remain limited to categories/sources of training data?

d) What legal framework can balance copyright holders' informational interests with AI innovation, trade secrets and technological development?

1.5 Scope and Methodology

The study adopts a doctrinal legal research methodology supported by a comparative approach. It examines primary sources including statutes, regulations, judicial decisions and official regulatory materials, along with secondary sources such as academic scholarship and policy literature. The principal jurisdictions selected for comparison are India, the European Union and the United States, with additional jurisdictions considered only where they provide significant legal developments relevant to the emerging right to know in AI training data.

2 Conceptual and Legal Foundations of the Right to Know

2.1 Meaning and Scope of the Right to Know

The right to know means that a copyright owner is entitled to get enough information in order to assess whether and in what way their copyrighted work has been used in the training of AI systems. In order for this right to be of any value, the information provided should specify whether the work was used, where it was obtained from, the nature and extent of its use, and, where appropriate, the AI system or the developer responsible. This kind of information allows copyright owners to understand how their works have been treated and to assess their legal rights. At the same time, however, the extent of the information that has to be disclosed should be proportionate and must not unduly interfere with legitimate AI development, trade secrets or confidential information.[7]

2.2 The Copyright Owner's Informational Interests

Copyright holders have exclusive rights which can be influenced by the way their works are used in the training of AI systems, such as the right to reproduce, communicate or distribute where relevant, the right to adapt and the right to license. Knowledge of the data being used allows copyright owners to decide whether or not these rights have been exercised and whether or not further legal steps are needed. Such knowledge can help copyright owners in negotiations over licensing, in keeping an eye on unauthorised use, in claiming remuneration where this is legally possible, and in taking enforcement action or pursuing appropriate remedies. The right to know therefore serves to enable the practical exercise of the copyright rights which already exist rather than in itself creating a new substantive right; its importance consists in reducing the information gap between copyright owners and AI developers.[8]

[7] *Id.* recital 107.

[8] Copyright Act, 1957, § 14.

2.3 The Problem of Hidden Use and Copyright Infringement

Normal cases of copyright infringement usually revolve around a specific act involving a work that is the subject of copyright protection, allowing the copyright holder to identify and oppose the supposed infringing use. With AI training, however, this situation becomes more complicated since copyrighted works can be gathered, copied, processed and included in large datasets without any obvious interaction taking place with the copyright holder. This results in an "invisible use" problem, meaning that the owner may find it difficult to tell whether a given work was part of the training data or later used in the development of the AI model. Even if infringement has taken place, the lack of available information about the data sources and the processing makes it much more difficult for copyright owners to detect the infringement, to gather evidence and to take enforcement action.

2.4 The Right to Know as Compared with the Right to a Licence or Payment

The right to know must be separated from the right to obtain a licence or to be paid. The fact that information about the use of a copyrighted work is available does not mean that the copyright owner is entitled to payment or that the AI developer was legally obliged to get prior permission. On the other hand, transparency can act as a necessary condition for exercising the copyright and commercial rights which already exist. Knowing that a work has been used may allow the owner to decide whether or not licensing negotiations, claims for payment or actions for infringement are suitable in accordance with the relevant law. Hence, the right to know is mainly of an informational nature, whereas licensing and remuneration rely on the specific legal framework relating to the way in which the work has been used.

2.5 The Right to Know Takes the Form of Dataset Transparency

Dataset transparency offers the practical means by which the right to know can be put into practice in the context of AI. Since a copyright owner cannot properly ascertain whether their work has been used without having access to adequate information about the sources, categories or origin of the training data, transparency requirements may therefore necessitate that AI developers make available appropriate details regarding the datasets, while taking into account confidentiality, security and legitimate commercial interests. The main argument of this paper is that the right to know only becomes practically meaningful in so far as legal mechanisms require a certain level of transparency regarding the training data used in AI. This kind of transparency does not have to involve the disclosure of each individual work but should instead provide sufficient information to enable the meaningful verification and exercise of copyright interests.

3 COPYRIGHT, AI TRAINING AND THE EMERGING TRANSPARENCY FRAMEWORK: A COMPARATIVE ANALYSIS

3.1 Copyright Protection and AI Training Data

The process of training AI usually involves gathering huge amounts of digital content by means such as web scraping, downloading and extracting data. In cases where the material obtained consists of works protected by copyright, it becomes necessary to consider whether the acts of copying, storing, processing and including those works in the training datasets constitute a breach of the copyright holders' exclusive

rights. When carrying out a legal assessment it is essential to make a distinction between the first collection of a work, its reproduction or processing while preparing the dataset, and its later use in training an AI model. The mere fact that copyrighted material is present in a dataset does not mean that infringement has taken place; instead, the legality of each particular action must be determined in accordance with the relevant copyright rules and any statutory exceptions. This provides the solid basis for looking at transparency: before a copyright owner can judge whether their rights have been affected, they might need to be given information regarding the collection and handling of their work.

3.2 European Union

The European Union has one of the most advanced sets of regulations regarding copyright and transparency in relation to data used for training AI. The Directive on Copyright in the Digital Single Market includes specific rules about text and data mining (TDM). According to Article 3, eligible research organisations and cultural heritage institutions may carry out reproductions and extracts for scientific research purposes, provided that certain conditions are met. Article 4 introduces a wider exception for TDM, while at the same time enabling copyright holders to retain their rights in respect of works that have been made available online using suitable methods.

The EU framework is especially important in the light of the EU AI Act, which imposes on providers of general-purpose AI models duties relating to transparency. Such providers must keep and make available information about the content which has been used to train their models, in line with the applicable regulatory framework. The aim of these obligations is to enhance transparency regarding the training content while at the same time taking into account both copyright protection and the practical circumstances of large-scale AI development. The main question that this paper addresses is whether these obligations constitute an effective "right to know" for individual copyright holders. A general disclosure of the training content may contribute to greater transparency without necessarily enabling an individual author to determine whether a particular work was used. The EU framework thus offers a key opportunity to assess whether regulatory transparency can be turned into a real informational right.

3.3 United States

In the United States, the way in which copyright has addressed its relationship with AI training has mostly been carried out by means of copyright doctrine, fair-use analysis and litigation, not through a general legal right that would compel AI developers to make disclosure of their training data to copyright holders. When applying the fair-use framework, account must be taken of various factors such as the purpose and character of the use, the nature of the copyrighted work, the amount used and the effect on the potential market.[9]

The recent legal actions relating to AI training have led to doubts about whether copying copyrighted works for the purpose of machine learning can be considered fair use and whether such use is sufficiently transformative. Yet there are considerable difficulties in establishing the relevant facts. In order to determine whether specific works were actually used, copyright holders may need details about the training datasets, how the data was acquired, the filtering processes and the development of the model. This kind of information can be obtained through litigation and the discovery process, but this is quite different

[9] 17 U.S.C. § 107 (2018).

from having a general right granted by statute to obtain the information prior to starting any legal action.[10] The approach taken in the United States thus shows a clear difference between substantive copyright protection and the right to access information. The fair-use doctrine decides whether a given use may be legally allowed, while litigation and discovery offer the means of acquiring evidence. Neither of these, however, gives rise to a general and independent right to find out whether a particular copyrighted work was used in the training of AI systems.

3.4 International Developments

The United Kingdom, Japan and Singapore are examples of jurisdictions which offer helpful additional insights into AI training and copyright. These countries show that it is possible for different combinations of copyright exceptions, text-and-data-mining provisions, licensing arrangements and regulatory measures to be used in dealing with AI training. Such developments are relevant since they prove that transparency does not have to take only one legal form.

Nevertheless, these jurisdictions should still be treated as merely providing comparative examples rather than as the subject of separate country studies. The aim is to highlight various approaches which may shed light on certain aspects of the right to know, for instance those relating to the disclosure of data sources, the allowed use of computational methods, licensing arrangements or the limitations on copyright enforcement. Inclusion of such cases should therefore be restricted to those developments which make a direct contribution to answering the main comparative question – whether copyright owners can obtain meaningful information about the way their works are used in the training of AI systems.[11]

3.5 Comparative Analysis

The analysis comparing these approaches shows three main kinds of regulation. The European Union mainly follows a regulatory-transparency model, linking the copyright rules relating to text and data mining with specific duties imposed on general-purpose AI models. While this method places the greatest emphasis in law on transparency regarding the training data, there are still uncertainties as to whether general disclosure enables individual copyright holders to check that particular works have been used. In contrast, the United States mainly depends on litigation, the fair-use doctrine and evidentiary procedures. This way of proceeding lets the courts decide on a case-by-case basis whether it is legal to use data for AI training, but it offers no general statutory right to copyright owners to get information about how their data has been used; access to the relevant information may therefore have to be obtained through litigation and the discovery process.[12]

Other legal systems show examples of emerging or sector-specific strategies, usually achieving a balance between promoting AI innovation and protecting copyright by means of exceptions, licensing schemes or limited transparency measures. The comparison thus indicates that transparency and the right to know are related yet conceptually different. The EU shows how regulatory transparency can act as an institutional means of providing disclosure, whereas the United States illustrates the shortcomings of depending mainly on litigation to reveal details about hidden AI training practices. The main issue under

[10] Authors Guild v. Google, Inc., 804 F.3d 202, 214–25 (2d Cir. 2015).
[11] Copyright, Designs and Patents Act 1988, c. 48, § 29A (UK); Chosakukenhō [Copyright Act], Law No. 48 of 1970, art. 30-4 (Japan); Copyright Act 2021, No. 22 of 2021, §§ 243–244 (Sing.).
[12] Thomson Reuters Enter. Centre GmbH v. Ross Intelligence Inc., No. 1:20-cv-00613, 2025 WL 458520 (D. Del. Feb. 11, 2025).

comparison is therefore not merely which jurisdiction offers greater transparency, but whether the available mechanisms give copyright holders access to information that is specific enough, easy to obtain and practical enough to enable them to determine if their works have been used and to exercise their current legal rights.

4 THE INDIAN LEGAL POSITION AND THE PROPOSED RIGHT TO KNOW

4.1 Indian Copyright Act, 1957 and AI Training

The Indian Copyright Act, 1957 does not contain a provision specifically regulating the use of copyrighted works for training generative artificial intelligence systems. Nevertheless, the existing copyright framework provides the starting point for analysing whether activities associated with AI training may engage copyright protection. Section 14 confers upon copyright owners exclusive rights in relation to protected works, including rights concerning reproduction, adaptation and communication of the work to the public, subject to the limitations and exceptions contained in the Act. The relevance of these rights to AI training depends upon the particular stage of the technological process under consideration.[13]

AI training may involve several distinct activities, including collecting publicly available works, downloading or storing copies, preprocessing and extracting information, converting material into machine-readable formats, incorporating the processed material into training datasets and subsequently using that dataset to train an AI model. These activities should not automatically be treated as one legally indistinguishable act. In particular, the question whether the temporary or permanent storage of a copyrighted work constitutes reproduction must be examined separately from the question whether the subsequent computational use of that work falls within an applicable statutory exception.[14]

The issue has acquired particular significance in *ANI Media Pvt. Ltd. v. OpenAI OpCo LLC*, decided by the Delhi High Court in July 2026.[15] The proceedings involved allegations that copyrighted ANI material had been scraped and stored for training OpenAI's large language models. The Court considered, among other matters, whether storage of copyrighted material for AI training could constitute infringement and whether the Indian Copyright Act could apply where training and storage occurred on servers outside India.

The Court, at the prima facie stage, rejected the proposition that the location of the training servers automatically removed the dispute from Indian jurisdiction. It considered the process holistically, including the access to and transmission of data from India and the connection between training and outputs generated within India. This decision demonstrates that AI training presents questions that cannot be resolved simply by identifying where the final model is hosted. The geographical distribution of data collection, transmission, storage, training and output generation may make conventional territorial assumptions increasingly difficult to apply. At the same time, the case demonstrates why training and output should remain analytically distinct. A model may be trained on copyrighted material without necessarily reproducing that material in its outputs. Conversely, an output may reproduce a protected expression even where the legal basis of the underlying training activity remains contested. The copyright analysis must therefore identify the particular act allegedly engaging the owner's rights. The Indian framework

[13] Copyright Act, 1957, § 14.
[14] *Id.* §§ 14(a)(i), 52(1)(a).
[15] ANI Media Pvt. Ltd. v. OpenAI OpCo LLC, CS(COMM) No. 1028 of 2024 (Del. HC July 24, 2026).

consequently provides substantive copyright protection but does not expressly answer the preliminary informational question central to this paper, *"how can a copyright owner determine whether a particular work was collected, stored or used in AI training?"*

4.2 Fair Dealing and AI Training

Section 52 of the Copyright Act provides statutory exceptions to infringement. Section 52(1)(a), in particular, recognises specified acts of fair dealing in relation to literary, dramatic, musical and artistic works, including fair dealing for purposes such as private or personal use, including research, criticism or review and reporting current events, subject to the statutory framework. The application of this provision to AI training is legally significant because AI developers may argue that training constitutes a form of research or computational analysis. However, the Indian concept of fair dealing should not simply be equated with the broader American doctrine of fair use. Section 52 operates through specifically identified statutory purposes, and Indian courts have developed contextual approaches to determining whether a particular dealing falls within the relevant exception.

The *ANI v. OpenAI* proceedings provide an important contemporary development. The Court considered whether the storage of ANI's literary works for training the LLM underlying ChatGPT could fall within the expression "private or personal use, including research" in Section 52(1)(a)(i). At the prima facie stage, the Court accepted that AI-based model training could fall within the research component of the provision and concluded that the purpose requirement was satisfied.

Importantly, however, satisfying the purpose requirement did not automatically resolve the fair-dealing question. The Court separately considered whether the dealing itself was fair. It observed that there is no single uniform test equivalent to the American four-factor fair-use test and emphasised that fairness depends upon the facts, degree and overall circumstances of the particular use. Among the considerations identified was whether the training activity prejudiced the copyright owner's interests, including actual or potential commercial harm. The Court ultimately held, at the interim stage, that OpenAI's storage of ANI's literary works for training fell within Section 52(1)(a) and therefore did not amount to infringement under Section 51 on the facts before it. It also found that ANI had not established, at that stage, substantial reproduction or memorisation of its works in the outputs relied upon. The Court expressly stated that its observations were for the purpose of deciding the interim application and would not determine the final outcome of the suit.

This is particularly important for the present research. The decision should not be interpreted as establishing that all AI training in India is automatically protected by fair dealing. Rather, it demonstrates that the legality of training may depend upon the purpose, nature of the use, commercial context, potential prejudice to the copyright owner and other factual circumstances. It also reinforces the importance of transparency. A copyright owner cannot effectively evaluate commercial prejudice, potential infringement or the fairness of a particular use without knowing whether and how the work was incorporated into the training process.

4.3 Absence of an AI-Specific Transparency Framework

Although Indian copyright law provides substantive rights and statutory exceptions, it does not presently establish a comprehensive, AI-specific transparency regime giving copyright owners a clear statutory mechanism to determine whether their individual works have been used in AI training.

There is no general copyright provision expressly requiring an AI developer to provide an individual copyright owner with:

a) confirmation that a particular work was used for training;
b) identification of the dataset in which the work appeared;
c) information concerning the source from which the work was obtained;
d) information regarding whether the work was actually used during model training;
e) identification of the AI model trained using the work; or
f) a dedicated procedure through which the copyright owner can verify or challenge such use.

This distinction is fundamental. Copyright protection and copyright transparency are not the same thing. Section 14 may establish the rights of the copyright owner, but it does not necessarily provide the information required to determine whether those rights have been engaged. The absence of a dedicated transparency mechanism becomes particularly problematic because AI training datasets can be extremely large and may be compiled from multiple sources. A copyright owner may discover that an AI system produces content resembling their work, but such similarity alone may not establish that the particular work was included in the training dataset. Conversely, the absence of visible reproduction in an output does not establish that the work was never used during training.[16]

The *ANI* proceedings illustrate this evidentiary complexity. The Court noted that although the storage of ANI's original literary works during training was admitted, the plaintiff had not provided sufficient factual material to establish the alleged memorisation and regurgitation of those works in the outputs relied upon. The Court treated the question of memorisation and regurgitation as requiring evidence. This reveals an important gap between knowing that AI training occurred and knowing precisely what copyrighted works were used in that training.

Accordingly, India's current framework may address the substantive question of infringement through existing copyright principles, but it does not provide an equally developed answer to the preceding informational question.

4.4 Evidentiary Challenges in Proving AI Training Data Use

The most significant difficulty for copyright owners may therefore be evidentiary rather than purely substantive. Traditional copyright disputes generally involve an identifiable work and an identifiable act of copying, communication, adaptation or other restricted use. AI training creates a different evidentiary environment. The allegedly relevant copying may occur during automated data collection and preprocessing, potentially involving enormous quantities of works. The copyright owner may have no direct observation of the process and may have no access to the underlying technical records. This produces a significant information asymmetry.

The copyright owner may possess:

- the original copyrighted work;
- evidence concerning authorship and ownership; and
- evidence that the work was publicly available.

The AI developer, however, may possess:

[16] DEP'T FOR PROMOTION OF INDUS. & INTERNAL TRADE, WORKING PAPER ON THE INTERFACE BETWEEN ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW, PART I (Gov't of India 2025).

- data acquisition records;
- dataset documentation;
- source lists;
- filtering records;
- preprocessing information;
- provenance information;
- training logs; and
- information concerning the particular model and training stage.

The person seeking to enforce copyright may therefore need evidence that is primarily controlled by the alleged infringer. This raises questions concerning the burden of proof and access to evidence. A copyright owner cannot reasonably be expected to prove an invisible technical process solely through speculation. At the same time, requiring AI developers to disclose unrestricted datasets could create serious concerns relating to trade secrets, confidential business information, cybersecurity and model security. A proportionate solution must therefore balance these competing interests.[17]

Digital and electronic evidence may be particularly important. Dataset records, server logs, acquisition records, metadata, provenance information and technical documentation could potentially assist a court in determining whether a particular work was collected or processed. However, the technical complexity of AI systems may make such evidence difficult for a conventional copyright litigant to obtain, interpret and challenge. The *ANI* decision demonstrates this evidentiary problem. The Court distinguished between the existence of stored training material and the claim that the model subsequently memorised and reproduced particular works. It held that the latter issue could not be established merely from the outputs relied upon at the interim stage.

This distinction is central to the proposed right to know. If the copyright owner cannot obtain information about training data use, the owner may be unable even to formulate the evidentiary basis of a potential infringement claim. Thus, transparency is not merely a policy preference. It may have direct implications for access to justice and the practical enforceability of copyright.

4.5 Can an Emerging Right to Know Be Derived from Existing Indian Law?

The absence of an express statutory right does not necessarily mean that the concept has no place within Indian law. A potential right to know could develop indirectly through existing substantive rights, procedural mechanisms, judicial interpretation, contractual arrangements and future regulatory intervention.

- First, existing copyright rights provide the substantive foundation. If a copyright owner possesses exclusive rights over reproduction, adaptation and other protected acts, information concerning the exercise of those rights may become practically necessary for enforcement. The right to know could therefore be conceptualised as an informational mechanism supporting existing rights rather than as an entirely new copyright entitlement.
- Second, procedural and evidentiary mechanisms may provide a route for obtaining relevant information in litigation. Courts may, depending upon the applicable procedural framework and facts, require parties to produce relevant evidence. Such mechanisms could help address information

[17] U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING (2025).

asymmetry once litigation has commenced. Their limitation, however, is that they may provide information only after the dispute has reached a sufficiently advanced procedural stage.

- Third, judicial interpretation may gradually develop principles concerning disclosure where AI training and copyright disputes come before Indian courts. *ANI v. OpenAI* is particularly important in this respect because it demonstrates judicial engagement with training data, territorial jurisdiction, storage, fair dealing and the evidentiary distinction between training and output.
- Fourth, regulatory intervention could create a more systematic mechanism. Instead of requiring every copyright owner to initiate litigation to obtain information, legislation or subordinate regulation could impose proportionate disclosure obligations upon relevant AI developers.
- Fifth, contractual and licensing mechanisms could establish transparency obligations voluntarily. Where AI developers obtain training data through licensing arrangements, contracts could require reporting, attribution, provenance documentation or payment mechanisms.[18]

These possibilities demonstrate that the right to know could emerge incrementally. However, none presently establishes a comprehensive and universally applicable Indian statutory right enabling copyright owners to determine whether their works have been used in AI training. The more defensible position, therefore, is that Indian law presently provides the foundations for an informational interest but does not clearly recognise a standalone AI training "right to know."

4.6 Need for Legislative Recognition

The preceding analysis indicates that relying entirely upon existing copyright law may leave a significant practical gap. Copyright owners may possess substantive rights but lack the information necessary to determine whether those rights have been affected.

A future Indian framework should therefore consider recognising a proportionate transparency obligation concerning AI training data. Such a framework need not require disclosure of every individual item contained within a massive dataset. Instead, it could establish graduated obligations depending upon the nature and scale of the AI system.

Possible requirements could include disclosure of:

- significant categories and sources of training data;
- information concerning publicly accessible datasets;
- mechanisms for recording data provenance;
- information sufficient to verify the use of a particular copyrighted work where a credible claim is raised;
- procedures for copyright owners to submit verification requests; and
- appropriate documentation for regulatory or judicial review.

Such obligations should be balanced against legitimate interests in trade secrets, confidential information, cybersecurity and protection against misuse of technical information. The objective should therefore not be unrestricted access to proprietary datasets but sufficient information to make copyright protection practically meaningful. India could consequently adopt a model in which the right to know operates as an informational and procedural entitlement rather than as an automatic right to prohibit AI training or demand remuneration. Once use is established, the copyright owner could then determine

[18] Regulation 2024/1689, recital 107, 2024 O.J. (L 1689) 1 (EU).

whether existing copyright law provides grounds for licensing, remuneration, enforcement or other remedies. The Indian position therefore reveals the central argument of this paper: the effectiveness of copyright protection in the AI environment depends not only upon the substantive scope of copyright but also upon the availability of information concerning the use of copyrighted works. The absence of a dedicated transparency mechanism may leave copyright owners facing an evidentiary disadvantage that existing copyright provisions were not designed to address.[19]

5 CONCLUSION

The increasing use of copyrighted works in AI training has exposed an important limitation in traditional copyright frameworks: the effectiveness of a legal right may depend upon the ability of its holder to obtain information concerning its use. Copyright owners may possess exclusive rights over their works, yet remain unable to determine whether those works have been collected, incorporated into training datasets or used to develop particular AI systems. The resulting information asymmetry creates difficulties not only for infringement claims but also for licensing negotiations, remuneration and other legitimate exercises of copyright interests.

The comparative analysis demonstrates that jurisdictions are responding to this problem through different legal approaches. The European Union has moved towards a regulatory model that places greater emphasis on transparency concerning AI training content, while the United States continues to rely substantially upon fair-use doctrine, litigation and evidentiary mechanisms. Other jurisdictions have adopted more limited or sector-specific approaches. These differences demonstrate that transparency can be structured through different legal mechanisms and need not necessarily require unrestricted disclosure of proprietary datasets.

For India, the existing Copyright Act, 1957 provides a substantive framework through rights of reproduction, adaptation and communication and statutory exceptions under Section 52. However, it does not expressly establish a comprehensive mechanism through which copyright owners can determine whether their works have been used in AI training. Recent judicial engagement with AI-training disputes demonstrates that existing copyright principles can address particular questions of legality, but the evidentiary problem concerning hidden training-data use remains significant. This paper therefore proposes the recognition of a legally enforceable but proportionate right to know. Such a right should not automatically create an entitlement to licensing, remuneration or prohibition of AI training. Instead, it should provide copyright owners with sufficient information to determine whether their works have been used and to assess what legal or commercial action may be appropriate.

A three-level transparency model is proposed. General transparency would require disclosure concerning categories and major sources of training data and relevant copyright policies. Targeted disclosure would permit a copyright owner with a credible basis for suspicion to seek verification of whether a particular work was used. Evidentiary disclosure would operate during formal proceedings under judicial supervision, with appropriate safeguards for confidential information and trade secrets. Ultimately, the right to know should be understood as a precondition for meaningful enforcement rather than an independent right to control AI development. Where there is no reasonable means of knowing whether a copyrighted work has been used, the practical value of existing copyright protection may be substantially

[19] *Id.*

weakened. A balanced Indian framework should therefore reconcile the legitimate interests of copyright holders with innovation, confidentiality and technological development. Meaningful AI regulation requires not unrestricted disclosure, but sufficient, proportionate and actionable transparency.

* * *

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.11

CHAPTER 11

Copyright Infringement and Corporate Liability: Navigating the Modern Digital Landscape

DR. PALLAVI PAL*

The evolution of Artificial Intelligence (AI) has revolutionized the production of literary, artistic, musical, and digital content, prompting a reassessment of established copyright doctrines concerning authorship, ownership, and infringement. The technological systems are such that they are capable of independently creating content with no human intervention, or minimal intervention resulting in arising of complex legal questions concerning authorship, ownership, originality, and liability. The traditional copyright laws were primarily designed to protect human creativity and therefore struggle to address the legal status of AI-generated works. This has given rise to the new issues relating to AI developers and operators, whether they may be held liable for direct copyright infringement where AI-generated outputs substantially reproduce or closely resemble protected copyrighted works. This issue has assumed considerable importance in light of the rapidly advancing capabilities of generative AI systems, such as ChatGPT and DALL·E, which are capable of producing sophisticated textual, visual, and audiovisual content that may incorporate recognizable elements of copyrighted material. Under prevailing statutory provisions and judicial precedents, AI companies may, in certain circumstances, be exposed to direct copyright liability for infringing outputs generated by their systems. However, the application of traditional copyright doctrines to AI-generated content presents novel legal complexities that challenge established principles of infringement analysis. Existing tests for determining liability were developed in the context of human authorship and conduct and therefore require substantial adaptation to address the distinctive characteristics of AI technologies. This paper examines three interrelated dimensions of this emerging legal challenge. Firstly, it analyses the foundational principles of copyright law governing corporate liability, with particular emphasis on the doctrines of direct, contributory, and vicarious infringement as applied to AI-generated content. The second part of the paper evaluates the applicability of these traditional doctrines to generative AI systems, including issues relating to the use of copyrighted training data, substantial similarity, and the doctrine of transformative use. The final part of the paper assesses the adequacy of existing legal frameworks in addressing challenges posed by AI-driven copyright disputes and proposes potential regulatory and judicial responses aimed at achieving an appropriate balance between the protection of intellectual property rights and the encouragement of technological innovation.

1 INTRODUCTION

Copyright law was built to protect creators against unauthorized copying, but digital technologies have made copying, distribution, and modification faster, cheaper, and harder to police. For corporations, this creates legal exposure not only when they directly use infringing content, but also when their platforms, employees, contractors, users, or algorithms facilitate infringement. The result is a legal environment in which companies must manage both intellectual property compliance and operational risk.

* Independent Practitioner and Academician.

The issue is especially pressing for platforms hosting user-generated content and for AI firms training or generating content using large datasets. In these settings, liability questions often turn on knowledge, control, financial benefit, and response speed once infringement is discovered.[1]

Copyright law, therefore, has historically recognized human creators as authors and vested them with a bundle of exclusive rights over their original works. These rights encompass the authority to reproduce, publish, distribute, adapt, and commercially exploit their creations, derive economic benefits therefrom, transfer or assign copyright ownership, and safeguard their honour and reputation against any distortion, mutilation, or misuse that may prejudice their interests. The legal recognition of authorship has thus served as the foundation upon which copyright protection and liability for infringement have been constructed.

At the international level, the Berne Convention for the Protection of Literary and Artistic Works serves as a cornerstone of copyright protection. The Convention guarantees international recognition and protection of literary and artistic works, extending its benefits not only to authors but also to their successors in title. By facilitating the enforcement of authors' rights and preserving the integrity of creative expressions, the Convention strengthens the legal position of authors across jurisdictions. Furthermore, it promotes transparency in copyright ownership by encouraging the identification of authors through their names or pseudonyms, thereby reducing uncertainty concerning authorship. Consequently, the Berne Convention occupies a pivotal role in the protection, recognition, and preservation of authors' contributions to literature and the arts, while providing the foundational framework for the global regulation of authorship and copyright.

Notwithstanding these international efforts, fundamental questions relating to the identification of authors and the attribution of copyright ownership have continued to attract significant scholarly and legal attention. The emergence of software-assisted creation has long generated debate within intellectual property jurisprudence, particularly regarding the role of computers and, more recently, artificial intelligence as instruments of creative production. Although technological capabilities have advanced considerably over the decades, the legal questions surrounding authorship and ownership of machine-assisted or machine-generated works remain unresolved, necessitating further judicial interpretation, legislative intervention, and doctrinal clarification.[2]

The rise of artificial intelligence has further complicated the copyright landscape by introducing concerns that extend beyond authorship and ownership to include copyright infringement and corporate liability. Modern generative AI systems are trained on vast datasets that frequently contain copyrighted literary, artistic, musical, and audiovisual works. This has generated significant legal controversy regarding whether the unauthorized use of copyrighted material during the training process constitutes copyright infringement. Equally contentious is the question of liability when AI-generated outputs substantially reproduce or closely resemble protected works. As AI systems increasingly operate with limited human intervention, determining responsibility for infringing acts becomes increasingly complex. Courts and scholars have debated whether liability should rest with the developers of AI models, the corporations that deploy and commercialize such systems, the users who generate the outputs, or a combination

[1] For cases in which courts have found companies secondarily liable for their users' direct copyright infringement, *see* Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 935–36 (2005); A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004, 1024 (9th Cir. 2001).

[2] Pamela Samuelson, *AI Authorship?*, 63 COMMC'NS ACM 20, 22 (2020).

thereof. Consequently, the issue of corporate liability has emerged as a central concern in contemporary copyright discourse, particularly as technology companies continue to profit from AI-generated content while relying upon copyrighted materials to develop and refine their models.

Recognizing the challenges posed by technological innovation, the United States established the Commission on New Technological Uses of Copyrighted Works (CONTU) in 1974. The Commission examined the implications of computer technology for copyright law and recommended significant reforms, including the recognition of computer programs as copyrightable subject matter. It further advocated extending copyright protection to encompass the various uses of software in computer-related operations. Importantly, CONTU emphasized that lawful owners of computer program copies should retain the freedom to use and adapt such programs within the limits prescribed by copyright law, thereby balancing the interests of copyright holders with those of users.

Subsequently, in 1986, the United States Office of Technology Assessment (OTA) undertook a comprehensive study examining the impact of emerging technologies on intellectual property rights. The OTA evaluated the adequacy of the federal copyright framework in addressing technological innovation and concluded that rapid technological developments were exerting a transformative effect on the intellectual property system. The study underscored the necessity of continual legal adaptation to ensure that copyright law remains responsive and effective within an evolving technological landscape.

The advent of artificial intelligence has further intensified these debates by fundamentally transforming the creation of literary, musical, artistic, and digital works. AI systems are no longer confined to facilitating human creativity; they increasingly generate original content with minimal human intervention. A notable illustration of machine-generated creativity emerged in 2016 with the development of a Japanese artificial intelligence system that produced a novel entitled *The Day a Computer Writes a Novel*. The work attracted considerable attention and was shortlisted for the prestigious "Hoshi Shinichi Literary Award", demonstrating the growing sophistication of AI-generated creative outputs. Likewise, the release of "Hello World" by SKYGGE highlighted the expanding collaboration between artificial intelligence and human creators in the field of music. Widely regarded as one of the earliest examples of a fully AI-generated musical composition, the project exemplifies the transformative potential of artificial intelligence within creative industries.

Recent litigation against major artificial intelligence companies has brought these issues into the limelight. Authors, artists, musicians, publishers, and media organizations have initiated lawsuits alleging that AI developers unlawfully reproduced copyrighted works during the training of generative models and subsequently generated outputs that imitate or replicate protected expression. These disputes have raised critical legal questions concerning the scope of copyright infringement in the context of machine learning, the applicability of traditional doctrines such as fair use, and the extent to which corporations developing and deploying AI systems may be held directly, vicariously, or contributorily liable for infringing activities. As courts across multiple jurisdictions grapple with these novel challenges, the intersection of AI-generated creativity, copyright infringement, and corporate accountability has emerged as one of the most significant and rapidly evolving areas of contemporary intellectual property law.

The deployment of conventional copyright principles in the AI context has revealed significant gaps within the current legal framework. On the one hand, an overly expansive interpretation of infringement may impede technological innovation and the development of transformative AI systems. On the other hand, inadequate protection may undermine the legitimate interests of copyright holders and erode

incentives for creative production. The autonomous learning processes and content-generation capabilities of modern AI systems necessitate a more nuanced legal approach capable of reconciling intellectual property protection with the promotion of technological advancement.

In the light of the above stated scenario, it is important to emphasize that the paper contributes to the existing scholarship in several significant ways through its comparative study of jurisdictions operating under both civil law and common law systems. This paper examines three interrelated dimensions of this emerging legal challenge. *Firstly*, it analyses the foundational principles of copyright law governing corporate liability, with particular emphasis on the doctrines of direct, contributory, and vicarious infringement as applied to AI-generated content. *Secondly,* it evaluates the applicability of these traditional doctrines to generative AI systems, including issues relating to the use of copyrighted training data, substantial similarity, and the doctrine of fair use. *Finally*, it assesses the adequacy of existing legal frameworks in addressing contemporary challenges posed by AI-driven copyright disputes and proposes potential regulatory and judicial responses aimed at achieving an appropriate balance between the protection of intellectual property rights and the encouragement of technological innovation.

2 FOUNDATIONAL PRINCIPLES OF COPYRIGHT LAW GOVERNING CORPORATE LIABILITY

The rapid advancement of artificial intelligence technologies has challenged the traditional framework of copyright law by introducing novel questions concerning liability for infringing acts committed through autonomous or semi-autonomous systems. While copyright law was originally designed to regulate human conduct, the increasing involvement of corporations in the development, deployment, and commercialization of generative artificial intelligence has necessitated a re-examination of the principles governing copyright infringement and corporate accountability.

In this context, the doctrines of direct, contributory, and vicarious infringement serve as the primary legal mechanisms through which liability may be imposed upon corporate entities whose AI systems generate infringing content.

Direct copyright infringement occurs when a protected work is reproduced, distributed, publicly displayed, performed, or adapted without the authorization of the copyright owner. Traditionally, liability for direct infringement is imposed irrespective of the defendant's intention or knowledge, making it a strict liability offence in many jurisdictions. Therefore, in order to prove direct infringement, a plaintiff must establish ownership of a valid copyright and unauthorized copying of protected expression.

The application of this doctrine to generative AI presents significant challenges. AI systems are trained on extensive datasets that may contain copyrighted material, often collected from publicly accessible sources without explicit authorization from copyright owners. During the training process, copyrighted works may be reproduced, stored, and analysed by machine-learning algorithms. Consequently, rights holders have argued that such acts constitute unauthorized reproduction under copyright law. Furthermore, liability concerns arise when AI-generated outputs substantially resemble or replicate protected works, raising questions as to whether the resulting output constitutes an infringing reproduction.[3]

In the light of the context, a number of AI companies and scholars have argued that the AI training process constitutes "fair use" and is therefore non-infringing. "This doctrine serves as a critical limitation on the copyright holder's exclusive rights and provides potential defence for AI companies. The four

[3] U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING (2025).

testing parameters for finding out the fair use: the purpose and the character of the use, the nature of the copyrighted work, the amount used and the market effect."[4] While the courts have generally found that technological innovations transforming the copyrighted work to make a new work are considered as fair use,[5] the AI systems process and recombine the information in such a way that blurs the lines between inspiration, transformation and reproduction. Here it becomes difficult for the court to decide as to how to apply the 'fair use' doctrine.[6]

Furthermore, the market effect analysis opens up a new dimension in the AI context. Courts must consider not only the immediate impact on existing markets for copyrighted works but also the potential future markets that AI generated content might create or disrupt.[7]

In the context of corporate liability, AI developers and technology companies may face claims of direct infringement where they actively reproduce copyrighted material during model training or where the architecture of their systems predictably generates infringing outputs. However, determining direct liability becomes more complex where the alleged infringement results from autonomous computational processes rather than deliberate human action. This has led courts and scholars to debate whether traditional standards of direct infringement are adequate for addressing AI-generated content.

2.1 Contributory Copyright Infringement

Contributory infringement is a form of secondary liability that arises when a party knowingly contributes to, facilitates, or induces the infringing conduct of another. The doctrine emerged to prevent entities from avoiding liability by distancing themselves from the actual act of infringement while nonetheless enabling or encouraging such conduct.

To establish contributory infringement, two elements generally must be satisfied: first, the defendant must possess actual or constructive knowledge of the infringing activity; and second, the defendant must materially contribute to or induce the infringement. In the digital environment, courts have frequently applied this doctrine to online service providers, file-sharing platforms, and technology intermediaries that facilitate copyright violations by users.

The doctrine assumes particular significance in relation to generative AI systems. Technology companies that design, train, and distribute AI models may be accused of materially contributing to copyright infringement if their systems enable users to generate unauthorized reproductions of copyrighted works. For example, where a corporation is aware that users are employing its AI tools to generate content that closely imitates copyrighted material and fails to implement reasonable safeguards, it may be argued that the corporation has materially contributed to the infringing conduct. Similarly, the intentional promotion of AI systems as capable of reproducing particular artistic styles, literary works, or copyrighted content may strengthen claims of contributory liability.

The contributory infringement framework is especially relevant because it allows courts to assess corporate conduct even where the corporation itself is not the direct creator of the infringing output. As a result, it provides an important mechanism for addressing the intermediary role played by AI developers in the copyright ecosystem.

---

[4] 17 U.S.C. § 107 (enacted by the Copyright Act of 1976, Pub. L. No. 94-553, 90 Stat. 2541); *see also* H.R. REP. NO. 94-1476, at 65–66 (1976).

[5] *See* Authors Guild v. Google, Inc., 804 F.3d 202, 214–18 (2d Cir. 2015).

[6] *Id.* at 229.

[7] *Id.*

2.2 Vicarious Copyright Infringement

Vicarious infringement represents another form of secondary liability that does not require proof of knowledge or intent. Instead, liability is imposed where a defendant possesses the right and ability to supervise or control the infringing activity and derives a direct financial benefit from that activity. The doctrine is rooted in agency principles and reflects the notion that entities benefiting from infringement should bear responsibility when they possess the capacity to prevent it.

In the context of AI-generated content, vicarious liability may be particularly relevant to corporations that operate and monetize generative AI platforms. Technology companies often maintain substantial control over the design, training methodologies, content moderation systems, and deployment of AI models. They may also derive significant economic benefits through subscription fees, licensing arrangements, advertising revenue, or increased market valuation resulting from the popularity of their AI services.

Where a corporation possesses the technical capability to restrict infringing outputs yet fails to implement effective safeguards, courts may consider whether the company exercised sufficient control over the infringing activity. Likewise, where the availability of copyrighted content enhances the commercial value and attractiveness of an AI system, the financial benefit requirement may be satisfied. Consequently, vicarious liability offers copyright holders an avenue for holding corporations accountable even in circumstances where direct participation in the infringing act cannot be conclusively established.

The doctrines of direct, contributory, and vicarious infringement collectively provide the legal foundation for evaluating corporate responsibility in the age of generative artificial intelligence. However, their application to AI-generated content remains uncertain due to the unprecedented nature of autonomous machine learning systems. Unlike traditional copyright disputes involving identifiable human actors, AI-generated works often involve multiple stakeholders, including developers, model trainers, platform operators, and end users. This diffusion of responsibility complicates the attribution of liability and exposes limitations within existing copyright frameworks.

Recent litigation against major AI companies has highlighted these challenges by raising questions concerning unauthorized use of copyrighted works for model training, the generation of outputs that mimic protected expression, and the extent to which technology corporations should bear responsibility for the actions of their AI systems. As courts continue to grapple with these issues, the doctrines of direct, contributory, and vicarious infringement remain central to determining the contours of corporate liability in the emerging landscape of artificial intelligence and copyright law.

In light of these developments, it is increasingly evident that traditional copyright doctrines must either be adapted or supplemented through legislative reform to address the unique challenges posed by generative AI. Establishing clear standards of corporate accountability will be essential for balancing technological innovation with the protection of authors' rights and the integrity of the copyright system.

2.3 Human Authors v Non-Human Authors: A Comparative Analysis

Amid this backdrop, AI has caused global disruption within the copyright systems. The issues regarding the meaning and scope of "authorship" have been globally discussed. The Berne Convention has not

explicitly defined the term "author" but it laid down the broader perspective including both 'human'[8] (author) and 'legal entity'[9] (authorship) creators, meaning thereby the convention strongly refers to the individual who created the work.[10] Therefore, it could be said that copyright protection is exclusively granted to man-made creative works only.[11]

In India, the copyright regime attributes authorship in relation to computer-generated literary, dramatic, or artistic works to the individual who undertakes the necessary arrangements for the creation of such works. Nevertheless, notwithstanding the involvement of artificial intelligence in the creative process, the law continues to regard the human contributor as the primary creative force operating behind the generation of the work.[12]

The Copyright Act, 1957[13] does not expressly define the term "author" in the context of literary and artistic works. However, Section 2(d) of the Act provides a broad statutory interpretation of the concept. Under the provision, the author of a literary or dramatic work is identified as the person who creates the work, whereas in the case of a musical work, authorship is attributed to the composer. Similarly, for artistic works other than photographs, the "author" is recognized as the artist responsible for the creation. In the context of photographs, the Act specifically designates the person taking the photograph as the author.

The above-stated view is supported by the copyright systems in the European Union (EU). The legal framework governing copyright within the European Union comprises international treaties, Union legislation, and domestic laws enacted by individual Member States. The scope and nature of rights afforded to creators of original works differ across EU Member States, shaped by their respective legal traditions and national copyright regimes.

The notion of "authorship" under EU copyright law is therefore not entirely uniform and often depends upon the legal tradition followed within a particular Member State. Under the continental European approach, derived from the French concept of *droit d'auteur*, authorship incorporates both moral and economic rights. Moral rights such as the right of attribution and the right to preserve the integrity of a work are regarded as inalienable and cannot ordinarily be transferred or waived. Economic rights, however, are transferable in nature.

In contrast, common law jurisdictions, including the United Kingdom and, within the European Union, Ireland, Malta, and Cyprus, generally place greater emphasis on economic rights than on moral rights.[14] Although British and Irish copyright law recognize moral rights, such as the right to claim authorship and to object to derogatory modifications, these rights may be waived by the author. Consequently, the determination of who qualifies as an "author" under EU copyright law depends significantly

---

[8] Paris Act Relating to the Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, 3 U.N.T.S. 1,161 (entered into force Dec. 15, 1972).

[9] Caroline Anfray et al., *Reflection Paper on Copyright, Patient-Reported Outcome Instruments and Their Translations*, 16 HEALTH & QUALITY LIFE OUTCOMES 1 (2018).

[10] The wording of the above text strongly suggests that "author" and "authorship" for the purposes of the Convention refer to the individual who created the work. As a result, it seems acceptable to say that copyright protection is exclusively granted to man-made creative works. *See* Jane C. Ginsburg, *The Concept of Authorship in Comparative Copyright Law*, 52 DEPAUL L. REV. 1063, 1069 (2003).

[11] PAUL GOLDSTEIN, INTERNATIONAL COPYRIGHT: PRINCIPLES, LAW, AND PRACTICE (2001).

[12] Andres Guadamuz, *Artificial Intelligence and Copyright*, WIPO MAG. (2017), https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html.

[13] The Indian Copyright Act, 1957, § 2(d).

[14] COMPARATIVE LAW LIBRARY UNIT, EUROPEAN PARLIAMENTARY RESEARCH SERVICE, COPYRIGHT LAW IN THE EU: SALIENT FEATURES OF COPYRIGHT LAW ACROSS THE EU MEMBER STATES (2018), https://www.europarl.europa.eu/RegData/etudes/STUD/2018/625126/EPRS_STU(2018)625126_EN.pdf.

upon the applicable legal framework and the relative importance attributed to moral and economic rights within that system.[15]

Whereas, in the common law countries such as the US, the UK, Ireland, Malta, and New Zealand, there are discussions and debates regarding the recognition of "non-human" authors in copyright law. These jurisdictions adopt a predominantly *utilitarian approach*, wherein the primary objective of copyright law is to incentivize creativity and innovation while simultaneously promoting public access to creative works for the broader benefit of society.[16] Under this framework, comparatively lesser emphasis is placed upon the personality or identity of the author, thereby theoretically creating scope for discussions concerning non-human authorship. Nevertheless, judicial interpretation, particularly within the United States, has consistently clarified and limited the ambit of copyright protection to works originating from human intellectual effort and creativity.[17]

The Supreme Court of the United States, while interpreting the Copyright Act, has repeatedly affirmed that copyright subsists only in works created by human authors. "Historically, human authorship has been heavily stressed in US copyright law. The U.S. Copyright Act does not explicitly address AI-generated works, but the stance of the U.S. Copyright Office is particularly relevant."[18] The most recent edition of the Compendium of U.S. Copyright Office Practices states that, "The U.S. Copyright Office will register an original work of authorship, provided that the work was created by a human being."[19]

Copyright law in the United States is fundamentally grounded in utilitarian principles, whereby copyright protection functions as an incentive mechanism encouraging authors, artists, and inventors to create works for the broader benefit of society. This constitutional objective is reflected in Article I, Section 8, Clause 8 of the Constitution of the United States, which empowers Congress to promote the progress of science and the useful arts through the grant of exclusive rights.[20]

As previously discussed, United States copyright jurisprudence requires a sufficient degree of human intellectual effort and creativity in order to establish the originality necessary for copyright protection. Although Title 17 U.S.C. § 102 does not expressly address the copyrightability of AI-generated works, the United States Copyright Office has consistently clarified that works generated solely through artificial intelligence are not eligible for copyright protection unless they embody meaningful human authorship or creative contribution.[21]

It is, however, pertinent to note that the "work made for hire" doctrine under the amended Copyright Act of 1976 permits an employer or commissioning party to be recognized as both the legal author and owner of a work created by an employee within the scope of employment or pursuant to a contractual arrangement.[22] While this doctrine presupposes the existence of a human creator, it has nevertheless

---

[15] Copyright, Designs and Patents Act 1988, c. 48, §§ 77, 80 (UK), https://www.legislation.gov.uk/ukpga/1988/48/contents.
[16] *Id.*
[17] CHINA IPR SME HELPDESK, COPYRIGHT PROTECTION IN CHINA: A GUIDE FOR EUROPEAN SMES (2010), https://web.archive.org/web/20240926170458/https://intellectual-property-helpdesk.ec.europa.eu/system/files/2020-10/EN_Copyright_guide_Aug_2010.pdf.
[18] Anthi Gaidartzi & Irini Stamatoudi, *Authorship and Ownership Issues Raised by AI-Generated Works: A Comparative Analysis*, 14 LAWS 57, at 3 (2025), https://doi.org/10.3390/laws14040057.
[19] U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES § 306 (3d ed. 2021).
[20] U.S. CONST. art. I, § 8, cl. 8.
[21] *See* Anthi Gaidartzi & Irini Stamatoudi, *Authorship and Ownership Issues Raised by AI-Generated Works: A Comparative Analysis*, 14 LAWS 57, at 6 (2025), https://doi.org/10.3390/laws14040057.
[22] U.S. COPYRIGHT OFFICE, COPYRIGHT LAW OF THE UNITED STATES AND RELATED LAWS CONTAINED IN TITLE 17 OF THE UNITED STATES CODE (2022), https://www.copyright.gov/title17/title17.pdf.

generated scholarly debate regarding whether AI-generated outputs produced during the course of employment may indirectly qualify for protection under such arrangements. Nonetheless, this proposition remains legally unsettled and has yet to receive definitive judicial determination.

This judicial approach significantly narrows the applicability of copyright protection by confining it to creations derived from human creativity and intellectual conception. A notable illustration is *Burrow-Giles Lithographic Co. v. Sarony*[23], wherein the Court defined an author as the "originator" or "maker" to whom a work owes its origin. In that case, photographs were recognized as embodying the original intellectual conceptions of the author. Likewise, in *Goldstein v. California*[24], the Court reaffirmed that an "author," in the constitutional sense, refers to an individual who writes, composes, or creates an original work, thereby emphasizing the requirement of human intellectual contribution.

These precedents collectively underscore that the concept of authorship is intrinsically linked to human creativity, intellectual labour, and originality. The jurisprudence establishes that authorship arises from the act of creation undertaken by a natural person exercising independent intellectual effort. Consequently, non-human entities are excluded from recognition as authors within the constitutional and statutory framework of copyright law.

The requirement of originality was further elaborated in *Feist Publications, Inc. v. Rural Telephone Service Co.*[25], where the Court held that copyright protection extends only to works that are original to the author. The judgment clarified that a work must exhibit a sufficient degree of creativity and independent expression attributable to the author in order to qualify for copyright protection. Accordingly, mere compilations of facts or data lacking creative selection, coordination, or arrangement fail to satisfy the threshold of originality required under copyright law. The ruling thus reaffirmed that copyright subsists only in expressions reflecting distinct and independently created intellectual effort.

The settled jurisprudence of the Supreme Court emphasizing human authorship has consistently been reaffirmed by lower federal courts. These courts have uniformly rejected attempts to extend authorship recognition to non-human entities. In *Urantia Foundation v. Maaherra*,[26] the Ninth Circuit Court of Appeals held that a book allegedly authored by non-human spiritual beings could attract copyright protection only where there existed substantial human selection, arrangement, and editorial contribution. The Court thereby emphasized that meaningful human participation in the creative process remains indispensable for copyright protection.

Similarly, in *Cetacean Community v. Bush,*[27] the Court addressed questions concerning the legal standing of non-human entities in the context of federal environmental statutes. Although constitutional standing under Article III was discussed, the Court ultimately held that animals lacked statutory standing under those statutes, a rule the Ninth Circuit later applied to the federal Copyright Act in *Naruto v. Slater.*[28] Together, the decisions highlighted the distinction between constitutional standing and statutory entitlement under copyright law, while reaffirming that non-human entities cannot be recognized as authors within the prevailing legal framework.

---

[23] Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53 (1884).
[24] Goldstein v. California, 412 U.S. 546 (1973).
[25] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991).
[26] Urantia Found. v. Maaherra, 114 F.3d 955, 964 (9th Cir. 1997).
[27] Cetacean Cmty. v. Bush, 249 F. Supp. 2d 1206 (D. Haw. 2003), *aff'd*, 386 F.3d 1169 (9th Cir. 2004).
[28] Naruto v. Slater, 888 F.3d 418 (9th Cir. 2018).

In light of the settled judicial precedents and the prevailing practices of the United States Copyright Office, it is apparent that independently generated AI-created works are not presently eligible for copyright registration or protection. The existing legal framework maintains that works produced solely through autonomous artificial intelligence systems do not satisfy the requirement of human authorship necessary for copyright subsistence.

The established jurisprudence, together with the administrative guidelines issued by the Copyright Office, consistently reaffirms the foundational principle that copyright law is intended to protect the originality, intellectual labour, and creative expression of human authors. Although artificial intelligence technologies have evolved considerably and are capable of generating sophisticated and commercially valuable outputs, the absence of direct and meaningful human involvement in the creative process raises substantial legal concerns regarding the applicability of copyright protection to such works.

3 AI AND HUMAN INVOLVEMENT: A PREREQUISITE FOR COPYRIGHT?

The issue of whether human intervention constitutes an indispensable prerequisite for copyright protection remains one of the most debated questions at the intersection of artificial intelligence and intellectual property law. "Broadly, AI-generated works may be categorized into two distinct classes: *first,* works produced with substantial human involvement, wherein artificial intelligence functions merely as a tool employed to achieve human-directed creative objectives; and *second*, works generated autonomously by AI systems with minimal or no meaningful human participation.

Works falling within the former category generally satisfy traditional copyright standards, particularly the requirement of originality, which has historically been understood as the product of human intellectual effort, skill, and judgment. In contrast, autonomously generated AI outputs raise far more intricate legal, philosophical, and ethical concerns relating to creativity, intentionality, accountability, and the very concept of authorship in the absence of human agency."[29]

Notwithstanding this apparent distinction, the boundary between these categories is often difficult to delineate in practice. Numerous AI-generated outputs arise through a combination of human prompts, instructions, supervision, and autonomous generative processes, thereby creating hybrid works that resist rigid binary classification. Consequently, legislative and judicial responses to this issue differ significantly across jurisdictions. While certain legal systems require a substantial degree of human creative contribution as a condition precedent for copyright protection, others adopt a comparatively broader or more flexible approach toward computer-generated works.[30]

The World Intellectual Property Organization (WIPO) addressed the concern in the paper titled "Conversation on Intellectual Property Policy and Artificial Intelligence", referring to the increasing use of AI in creating literary and artistic works.[31] Excluding AI-generated works from copyright protection would reinforce the doctrinal position that copyright law is premised upon the protection and encouragement of human creativity, rather than machine-generated output. Such an approach would underscore

[29] Kalin Hristov, *Artificial Intelligence and the Copyright Dilemma*, 57 IDEA: J. FRANKLIN PIERCE CTR. FOR INTELL. PROP. 431, 435–36 (2017).

[30] Niloufer Selvadurai & Rita Matulionyte, *Reconsidering Creativity: Copyright Protection for Works Generated Using Artificial Intelligence*, 15 J. INTELL. PROP. L. & PRAC. 536, 536–43 (2020).

[31] WIPO SECRETARIAT, REVISED ISSUES PAPER ON INTELLECTUAL PROPERTY POLICY AND ARTIFICIAL INTELLIGENCE, WIPO/IP/AI/2/GE/20/1 REV. ¶ 23 (May 21, 2020).

the principle that the copyright system is intended to incentivize, reward, and preserve the fruits of human intellectual effort.

"Conversely, extending copyright protection to AI-generated works would reflect a policy orientation favouring the dissemination and availability of creative output to the public, irrespective of whether such works originate from human authorship or machine generation. On this view, the copyright regime would be understood as prioritizing access to creative works over a strict distinction between human and artificial creativity."[32]

During the second and third Sessions of the WIPO Conversation on Intellectual Property and Artificial Intelligence, the issue was examined through two competing perspectives. Those opposing copyright protection for AI-generated works advanced several arguments in support of their position. They contended that the involvement of human input, when combined with the automated functioning of AI algorithms, is predominantly technical in nature and does not, by itself, justify the conferral of copyright protection.[33] They further emphasized that human authorship remains a settled and foundational principle of copyright law, whereas AI remains an evolving concept lacking sufficient doctrinal certainty.

In addition, the opponents maintained that works generated by AI should properly fall within the public domain.[34] They argued that extending copyright protection to the arrangement, selection, and creation of AI-assisted output would raise significant policy concerns and could unduly restrict the public domain. In their view, subjecting a wider range of AI-generated literary and artistic works to copyright protection would diminish public access to such creations.[35]

4 Whether AI-Generated Creations Fulfil the Threshold of Originality Required for Copyright Protection?

In essence, the concept of "originality" denotes that a work seeking copyright protection must emanate from an identifiable author or creator and embody a degree of independent intellectual creation reflective of the author's personal expression, rather than constituting a mere reproduction or wholly derivative work.

Under common law jurisprudence, originality was historically associated with the "*sweat of the bron*" doctrine, which emphasized the labour, skill, and effort expended in the creation of a work. However, in jurisdictions such as the United States, the standard of originality has undergone gradual transformation, and mere industrious effort or labour, in the absence of creativity, is no longer considered sufficient to establish originality or attract copyright protection. This modern approach aligns with the utilitarian philosophy underlying copyright law, which seeks to encourage human creativity and incentivize authors by granting legal protection to original intellectual expression. There have been repeated attempts to address the issue of whether the AI generated works are to be considered as original or not. The judiciary has significantly responded and clarified time and again but there has always been inconsistency between the common law countries and the civil law countries.

---

[32] *Id.*

[33] WIPO Secretariat, Summary of Second and Third Sessions, WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI), WIPO/IP/AI/3/GE/20/INF/5 ¶ 36 (Jan. 8, 2021).

[34] *Id.*

[35] *Id.* ¶ 39.

The decision in *Hollinrake v. Truswell*[36] further clarified the doctrine of originality within copyright jurisprudence. The Court held that originality does not depend upon the novelty or uniqueness of a work; rather, it is determined by whether the work originates from the author's own intellectual and creative effort. Accordingly, a work may be regarded as original provided that it is independently created and not copied from another source. The ruling thus reaffirmed that the essence of originality lies in authorship and independent creation, rather than in the existence of absolute novelty. By emphasizing the requirement of personal authorship and the absence of copying, the decision significantly shaped the understanding of originality within the jurisdiction of England and Wales.[37]

Similarly, in *University of London Press Ltd v. University Tutorial Press Ltd*,[38] the Court observed that a work may be considered original where it reflects a sufficient degree of skill, labour, and judgment and is not reproduced from an existing work. This interpretation of originality, grounded in independent creation and the absence of copying, became highly influential within United Kingdom copyright law. It is noteworthy, however, that the traditional "skill and labour" standard in the United Kingdom has subsequently been shaped and refined by the jurisprudence of the Court of Justice of the European Union (CJEU).[39]

Similarly, the courts in the United States have consistently recognized that originality requires at least a minimal degree of creative input. In *Burrow-Giles Lithographic Co. v. Sarony,*[40] the Court emphasized that copyrightability is contingent upon the presence of creative intellectual labour. Likewise, in *Time Inc. v. Bernard Geis Associates*, the District Court acknowledged that originality in photography may be reflected through elements of personality and creative judgment, including the selection of the subject, composition, framing, lighting, and timing of the photograph. "While recognizing the artistic and visual nature of photographic works, the Court refrained from evaluating the artistic merit or aesthetic quality of the work itself."[41]

Further clarification regarding the originality threshold was provided in *Baltimore Orioles, Inc. v. Major League Baseball Players Association*[42], where the United States Court of Appeals for the Seventh Circuit held that originality comprises two essential requirements: first, the work must originate independently from the author, and second, it must embody at least a minimal degree of creativity. The Court thereby distinguished between independent origin and creative expression, emphasizing that both elements are necessary to satisfy the originality requirement. "Independent origin" requires that the work emanate from the author's own intellectual effort rather than constituting a copy or imitation, whereas the requirement of "minimal creativity" recognizes that even a modest degree of creative contribution may suffice for copyright protection.

Likewise, in *Alfred Bell & Co. v. Catalda Fine Arts, Inc.,*[43] the Court of Appeals elaborated upon the "doctrine of originality" by holding that originality necessitates two essential conditions. *First,* the work must originate from the author through independent creation. *Second,* the work must contain at least a

---

[36] Hollinrake v. Truswell, [1894] 3 Ch. 420, 427 (Eng.).
[37] *See* Hafiz Gaffar & Saleh Albarashdi, *Copyright Protection for AI-Generated Works: Exploring Originality and Ownership in a Digital Landscape*, 15 ASIAN J. INT'L L. 23, 38 (2025), https://doi.org/10.1017/S2044251323000735.
[38] Univ. of London Press Ltd. v. Univ. Tutorial Press Ltd., [1916] 2 Ch. 601 (Eng.).
[39] Newspaper Licensing Agency Ltd. v. Meltwater Holding BV, [2011] EWCA (Civ) 890 (Eng.).
[40] Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53 (1884).
[41] Time Inc. v. Bernard Geis Assocs., 293 F. Supp. 130 (S.D.N.Y. 1968).
[42] Balt. Orioles, Inc. v. Major League Baseball Players Ass'n, 805 F.2d 663 (7th Cir. 1986).
[43] Alfred Bell & Co. v. Catalda Fine Arts, Inc., 191 F.2d 99 (2d Cir. 1951).

minimal degree of novelty or creative distinction. By articulating these principles, the Court clarified that copyright protection extends only to works reflecting independent intellectual effort and some degree of creative individuality, however slight, and not to mere reproductions or mechanical imitations of pre-existing works.

The common law countries such as the US and the UK prefer authorship over novelty while finding the threshold of originality. On the other hand, the civil law countries place less emphasis on "labour and skill" and more on inherent connection between the author and the creation. In these jurisdictions, moral rights play a crucial role including author's entitlement to be credited for their work and to maintain control over integrity.[44]

5 CONCLUSION

The global legal framework governing AI-generated works remains fragmented, with jurisdictions adopting divergent approaches to authorship and copyright protection. While the United States and the European Union continue to emphasize human authorship, jurisdictions such as China have adopted comparatively broader approaches accommodating AI-assisted creativity. This lack of uniformity creates significant legal uncertainty in the rapidly expanding market for AI-generated content.

The study argues that existing copyright regimes, grounded in human-centric notions of originality and authorship, are insufficient to address the complexities of autonomous AI-generated works. Accordingly, there is a growing need for a coherent international framework capable of balancing innovation, legal certainty, ethical accountability, and the protection of human creativity. Such a framework may include *sui generis* protection, compulsory licensing mechanisms, or revised standards of authorship recognizing collaborative human–AI creation.

This paper examined the issue of copyright protection for AI-generated works and identified several important findings. *First*, originality remains a fundamental requirement for copyright protection, requiring a work to reflect human creative choices and intellectual contribution. While fully autonomous AI-generated works may fail to satisfy this threshold, AI-assisted works involving meaningful human input may qualify for protection.

Second, the Court of Justice of the European Union (CJEU) has significantly influenced the interpretation of originality by emphasizing creative freedom and independent intellectual creation. *Third*, the study highlights the need for continued legal and policy discussions to address emerging issues relating to authorship, ownership, and attribution in AI-generated works.

Accordingly, policymakers and researchers should focus on developing a balanced and adaptable legal framework capable of accommodating technological innovation while safeguarding the principles of human creativity and copyright protection.

* * *

[44] *See* Hafiz Gaffar & Saleh Albarashdi, *Copyright Protection for AI-Generated Works: Exploring Originality and Ownership in a Digital Landscape*, 15 ASIAN J. INT'L L. 23, 39 (2025), https://doi.org/10.1017/S2044251323000735.

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.12

CHAPTER 12

Protecting Trademark in the Metaverse: The Emerging Threat of AI-Generated Counterfeits

A. VEDHA VALLI*

Generative artificial intelligence tools allow metaverse users to generate virtual wearables, avatar accessories and other branded goods, which often infringe upon trademark rights. These infringing goods, however, are not static counterfeits but dynamically generated knockoffs that rarely explicitly name the trademark owner and take place in novel digital environments. This paper analyses whether and how Indian trademark laws can tackle such emerging forms of infringement based on the doctrine of "use in the course of trade", the "identical or similar goods" jurisdictional anchor and the "likelihood of confusion" standard. It argues that the ability of Indian trademark law to address these concerns is limited by the very existence of doctrines and tests that seek to make sense of trademark infringement in virtual settings, as well as the lack of a holistic regulatory approach to the liability of metaverse platforms under the Trade Marks Act, 1999 and the Information Technology Act, 2000. By relying on recent Indian, EU, US and UK jurisprudence and early cases on trademarks and AI-generated counterfeits, the paper highlights the jurisdictional fragmentation in addressing generative AI infringement in India and proposes a unified approach to enforcing trademark rights against dynamically generated virtual counterfeits.

Keywords: trademark infringement, generative AI, metaverse, intermediary liability, virtual counterfeiting

1 INTRODUCTION

The term "metaverse" refers to an immersive, digital environment, where users can interact in real time through avatars, engaging in social, commercial and creative activity that imitates the real-world experience. Unlike the previous virtual platforms, the metaverse consists of technologies such as augmented and virtual reality, blockchain, Gen AI etc., allowing users to create, own, and transact within it. Platforms like Roblox, Decentraland, and The Sandbox exemplify this shift. It creates spaces where users can build, buy, and sell virtual goods, effectively imitating the real world commerce in a decentralised, user generated economy.[1]

It is this dual character of the metaverse which acts as both a social space and an economic market, that has drawn many luxury and fashion brands to venture out in virtual reality shops and it is precisely this same openness and user generative capability that has made the metaverse equally vulnerable to infringement.[2]

* Assistant Professor at Department of Intellectual Property Law, The Tamil Nadu Dr. Ambedkar Law University, Chennai, Tamil Nadu, India.

[1] Tambiama Madiega, Polona Car & Maria Niestadt, *Metaverse: Opportunities, Risks and Policy Implications* (European Parliamentary Research Service Briefing PE 733.557, June 2022), https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733557/EPRS_BRI(2022)733557_EN.pdf.

[2] Press Release, Prada Group, *LVMH, Prada Group and Cartier Come Together to Form the Aura Blockchain Consortium Supporting the First Global Blockchain Dedicated to the Luxury Industry* (Apr. 20, 2021), https://www.pradagroup.com/en/news-media/press-releases-documents/2021/21-04-20-aura-blockchain-consortium.html.

Metaverse platforms embed generative artificial intelligence tools that let users create wearables, avatar skins, and branded accessories through natural language prompts. Many brands are launching their own metaverse stores, expanding the brand in the virtual world. To name a prominent one, Gucci, one of the most famous names in high end fashion, didn't want to miss the opportunity of being an innovator in the fashion industry by investing in the metaverse. They began their investment in May 2021 when they opened Gucci Garden on Roblox, celebrating their 100th anniversary. During this event, users could purchase NFT collectibles. Later that year, they sold their Dionysus bag for 350,000 Robux, which equals about $4,115.[3] Their next step was investing in land on The Sandbox to expand its Gucci Vault, an online store. This virtual store will give shoppers a more immersive experience and develop their NFT market further. If they continue to gain popularity, their NFTs could become a $56 billion market by 2030.[4]

As official virtual storefronts demonstrated that branded digital wearables and accessories could command real economic value, the same open-platform tools that let Gucci and other houses build immersive stores also made it just as easy for anyone, without permission or licensing, to model, mint, and sell near identical virtual imitations of those goods, so counterfeiting migrated into the metaverse as a natural extension of its legitimate luxury market. Where conventional counterfeiting involves a fixed infringing product, manufactured once and distributed many times, generative virtual counterfeiting is produced anew at each point of use, frequently without the user ever naming the brand being copied.[5]

This generative shift in metaverse counterfeiting operates through two distinct pathways of infringement. In the first, a deliberate infringer, whether an individual creator or an operator of a virtual storefront, consciously designs and lists a virtual good that reproduces a brand's logo, trade dress, or distinctive product silhouette, mirroring the intent driven infringement familiar from conventional counterfeiting, only executed through 3D modelling or NFT-minting tools rather than physical manufacture.[6]

In the second, and more novel, pathway, the infringement is generated automatically by the platform itself in response to an ordinary user's natural language prompt, where a user asking a generative AI tool for "a designer style handbag" or "a luxury monogrammed jacket" may receive an output that reproduces protected trade dress or a recognisable brand aesthetic without the user ever naming, intending to copy, or even being aware of the brand being emulated. This second pathway unsettles the doctrinal assumptions on which trademark and design law were built, both of which presuppose an identifiable infringing actor who has made a deliberate choice to copy. Where the copying is instead a byproduct of a generative model's training data and the platform's own output design, liability becomes difficult to locate in either the prompting user or a single authorial infringer, raising the question of whether responsibility should shift toward the platform or the AI tool provider that shaped the generative output in the first place. These unintentional creations by generative AI could result in licensing market dilution.[7]

---

[3] *Gucci Hosts Virtual Exhibition in Online Game Roblox to Mark Centenary*, DEZEEN (May 20, 2021), https://www.dezeen.com/2021/05/20/gucci-garden-roblox-virtual-exhibition/.

[4] Barton Goldenberg, *Gucci's Metaverse Makeover: A Case Study in Fashion Innovation*, ISM (Dec. 19, 2023), https://ismguide.com/gucci-xr-metaverse-case-study/.

[5] Bernard Marr, *How Luxury Brands Are Making Money in the Metaverse*, FORBES (Jan. 19, 2022), https://www.forbes.com/sites/bernardmarr/2022/01/19/how-luxury-brands-are-making-money-in-the-metaverse/.

[6] Sneha Madan, *The Palika Bazaar of the Metaverse: Is the Trade Marks Act, 1999 Equipped to Regulate Digital Counterfeits?*, KHURANA & KHURANA (July 11, 2026), https://www.khuranaandkhurana.com/the-palika-bazaar-of-the-metaverse-is-the-trade-marks-act-1999.

[7] Reinhard Oertli, *Trademark Disputes in the Metaverse and Enforcement Challenges* (WIPO Webinar, Trademark Enforcement Issues in the Metaverse, WIPO/WEBINAR/RESPECTFOR IP/2023/1, Jan. 26, 2023), https://www.wipo.int/meetings/en/details.jsp?meeting_id=75574.

This shift is no longer speculative. In November 2025, the UK High Court delivered the first judgment anywhere to find trademark infringement arising from an AI model's own generative output, when Stable Diffusion was shown to reproduce Getty Images' watermark in synthetic images it had never been directly instructed to create.[8]

This paper asks whether India's trademark and intermediary liability framework, designed around a stable, identifiable act of "use" by a defendant, can respond to infringement that is generative, ephemeral, and machine scaled. It begins by tracing the shift from static, listing based counterfeiting to on demand virtual generation, grounded in a concrete scenario and in the scale of the market now at stake, and situating itself against the small but growing foreign scholarship already engaged with the problem. It proceeds to work through Indian trademark doctrine strain by strain, analysing the threshold requirement of "use," the tests for similarity and confusion, the separate doctrine of dilution, and design law as a further, equally insufficient regime, pointing out that each was built around a discrete, comparably fixed act of infringement that generative output does not supply. The analysis proceeds to address platform accountability under Section 79 of the Information Technology Act, including India's own recent executive advisories on generative AI, before setting this domestic picture against comparable developments in the European Union, United States, United Kingdom, and Singapore. Drawing upon this doctrinal and comparative foundation, the paper proposes a coordinated enforcement framework for India, outlines what that would mean in practice for rights holders, platforms, and regulators even before any reform, and concludes by acknowledging the limits of a doctrinal and comparative analysis of this kind.

2 METHODOLOGY

The methodology deployed in this research is doctrinal and comparative. The relevant statutes, namely the Trade Marks Act, 1999 and the Information Technology Act, 2000 are examined alongside the leading Indian authorities interpreting the same. These are then placed in comparative perspective with a set of carefully chosen foreign precedents from the jurisdictions of the European Union, the United States and the United Kingdom. The choice of these jurisdictions is informed by the fact that these have provided the most detailed engagement with the issue of liability for AI generated or platform mediated virtual content thus far. The comparison is made in order to single out nuances, such as the EU's approach to the use of virtual counterfeits based on their perceived impact on the consumer, which may have a degree of applicability in the Indian context. Finally, since the Indian case law remains entirely silent on the issue of generative virtual counterfeits, the paper extrapolates from the existing Indian case law on keyword and marketplace liability and dilution, identifying the point at which such reasoning breaks down in the face of generative content.

3 THE GENERATIVE SHIFT

3.1 From Static Knock-Offs to On Demand Generation

The metaverse, understood as a persistent and interconnected network of 3D environments navigated by avatars, has evolved from a niche gaming concept to a commercial space in which branded virtual

[8] Getty Images (US) Inc. v. Stability AI Ltd., [2025] EWHC 2863 (Ch).

goods, digital fashion and NFT backed assets hold real economic value. Counterfeiting in this environment takes several forms distinct from physical world infringement. First, unauthorized replication occurs when third parties replicate and sell virtual wearables, skins or accessories which copy registered trademarks or trade dress of established brands, often offered on in-platform marketplaces or third party NFT exchanges without any licensing arrangement. Second, "asset spoofing" includes minting NFTs visually replicating the metadata or imagery of a genuine branded item, exploiting the lack of verification by many marketplaces of the authenticity of the underlying intellectual property before listing. Third, avatar and storefront impersonation enables bad actors to erect stores which closely mimic a brand's presence, deceiving users into purchasing counterfeit digital goods or providing payment credentials in what amounts to a metaverse native phishing scheme. These practices are compounded by the pseudonymity of underlying blockchain transactions in many metaverse platforms, which defeats traditional enforcement mechanisms which rely on identifying an infringer's physical location or corporate identity, and by the jurisdictional ambiguity of platforms whose servers, users, and hosting entities may be distributed across multiple countries simultaneously.

Online counterfeiting, as documented in Indian enforcement literature, has largely concerned fixed listings. A seller uploads a product description and image, and that listing persists until removed. Generative virtual counterfeiting inverts this structure. The infringing asset does not exist until a user's prompt calls it into being, and no two generations need be identical. This defeats the assumption, implicit throughout trademark doctrine, that an infringing article is a discrete, comparably fixed thing that can be seized, delisted, or compared side-by-side with the registered mark.

Three episodes illustrate this trajectory. First, Hermès International v Rothschild[9], the first NFT trademark trial in the United States, found that a digital artist's "MetaBirkin" NFTs, AI assisted digital images evoking Hermès Birkin trade dress, infringed and diluted Hermès marks, with the jury rejecting a First Amendment defence because use was found explicitly misleading rather than genuine artistic commentary. Second, Nike Inc v StockX LLC[10], concerning NFTs marketed as "Vault" tokens bearing Nike's marks, settled in August 2025 before trial, leaving unresolved whether such tokens are themselves infringing "goods" or merely digital receipts for physical products. Third, and most significantly, Getty Images v Stability AI confirms that a generative model itself can be the direct source of trademark infringing output independent of any human curation of a specific listing.

The scale at which this problem is emerging is not trivial. Industry estimates put the global virtual goods market at approximately USD 132–150 billion in 2025–26, with virtual goods and NFTs identified as among the fastest growing revenue segments in the broader metaverse economy that itself is projected to grow at a compound annual rate in excess of 40 per cent through the remainder of the decade.[11] Separately, the OECD and the European Union Intellectual Property Office estimate that counterfeit

[9] Hermès Int'l v. Rothschild, No. 22-cv-384 (JSR), 2022 WL 4671520 (S.D.N.Y. Sept. 30, 2022).

[10] Nike, Inc. v. StockX LLC, No. 22-cv-983 (VEC) (S.D.N.Y. filed Feb. 3, 2022).

[11] Global Growth Insights, *Virtual Goods Market Size, Share, Growth, and Industry Analysis* (updated Feb. 5, 2026), https://www.globalgrowthinsights.com/market-reports/virtual-goods-market-116419; Mordor Intelligence, *Virtual Goods Market Size, Forecast Report, Drivers & Opportunities 2031*, https://www.mordorintelligence.com/industry-reports/virtual-goods-market; Technavio, *Metaverse Market Analysis, Size, and Forecast 2026-2030* (Apr. 2026), https://www.technavio.com/report/metaverse-market-industry-analysis (last visited Oct. 2, 2026).

and pirated goods accounted for approximately USD 467 billion in global trade in 2021.[12] These two curves are converging. The same brand equity that drives physical counterfeiting is now a generation prompt away from being reproduced in virtual form that no manual takedown process was designed to absorb. For India, this convergence arrives alongside the continuing challenge of counterfeit enforcement in physical markets, making early attention to the generative dimension a matter of pre-empting a known failure pattern rather than responding to an entirely novel one.

Consider a metaverse platform with a built-in generative wardrobe tool. A user, without naming any brand, prompts the tool for "a luxury quilted handbag in black leather with a gold chain strap for my avatar." The tool, trained on a large corpus of fashion imagery, returns an asset that closely tracks the trade dress of a specific well-known handbag. The user did not intend to infringe and may not even recognise the resemblance. The platform did not curate or list or advertise any specific infringing item and the asset itself may never be generated in exactly that form again. Yet the output is, on any ordinary understanding, a virtual counterfeit of a real brand's protected trade dress. None of Sections 29(1), (2), or (4) of the Trade Marks Act was drafted with this fact-pattern in mind, and none of the twenty-six Louboutin factors squarely focuses on a platform whose only "act" was to build and operate the generative tool itself.

The problem is even more acute for trade dresses and get-up marks than word-marks. A user would not have to type in the name "Hermes" or "Louboutin" to generate an image that incorporates elements of the plaintiff's protected trade dress; a prompt of the type "quilted flap handbag with an interlocking-C clasp" or "stiletto with a glossy red sole" or a combination of the two would be sufficient, given that the diffusion style image generating models learn associations between concepts and visual appearances. Not that this is merely a theoretical problem. It was precisely the red sole that the CJEU was asked to consider in the Louboutin v Amazon case when ruling on the scope of the "marketplace" liability, and the word based approach to infringement, ubiquitous in the Indian jurisprudence on keyword and meta tag use, would be entirely unhelpful in a case concerning visual similarity that never invokes brand names.

4 Related Work

The scholarship on trademark liability for generative AI output is only just beginning to emerge, and is almost entirely outside of India. UK practitioner commentary published ahead of INTA 2026[13] (and therefore presumably in advance of key developments in the law) has flagged that Section 10(4) of the UK Trade Marks Act 1994 provides no scenario that cleanly describes an AI generated output, and frames it as an "underappreciated risk" for AI platforms and intermediaries. Speaking of EU law, the question is how the concept of "use" under European trademark law should be interpreted for AI generated signs and how liability could be allocated between the user, the model developer, and the platform. US practitioner literature has separately observed that generative tools can produce brand confusing content "without any human intent to infringe" and that prompt filtering safeguards are emerging as an industry

[12] OECD & EUIPO, Mapping Global Trade in Fakes 2025: Global Trends and Enforcement Challenges (May 2025), https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/05/mapping-global-trade-in-fakes-2025_5c812e3c/94d3b29f-en.pdf.

[13] Oliver Fairhurst & Victoria Owrid, *Trade Mark Liability in the Age of Generative AI*, Lewis Silkin (Mar. 6, 2026), https://www.lewissilkin.com/en/insights/2026/03/06/trade-mark-liability-in-the-age-of-generative-ai.

response despite the lack of a settled doctrine requiring them.[14] None of this speaks to the Indian position, and none situates the problem within a jurisdiction where trademark and intermediary liability are governed by two separate and uncoordinated statutes rather than a single trademark code.

5 DOCTRINAL STRAIN: THE "USE" REQUIREMENT

Section 29 of the Trade Marks Act, 1999 conditions infringement on use, in the course of trade, of a mark identical or similar to a registered mark in relation to identical or similar goods or services, in a manner likely to cause confusion. Each element presupposes an identifiable actor whose conduct constitutes the "use."

The Delhi High Court's Division Bench, in Google LLC v DRS Logistics (P) Ltd[15] and the subsequent Google LLC v MakeMyTrip India Pvt Ltd[16] appeal, held that the invisible use of a registered mark as a keyword in a paid advertising programme constitutes "use in advertising" under the Trade Marks Act, extending the concept beyond visual display on a webpage. Crucially, however, the Court's reasoning still anchored liability to Google's own commercial conduct in selling and prioritising the keyword which is an identifiable, repeatable business practice, not a one-off machine generated output. A subsequent contempt proceeding in 2026 in DRS Logistics (P) Ltd. & Anr. v. Google India Pvt. Ltd. & Ors[17] confirmed that even this expanded "use" doctrine is read narrowly by the courts. Google was held not to be under a general obligation to proactively monitor all future advertisements for a plaintiff's marks, only to act on specific complaints, underscoring how far Indian doctrine remains from any duty to police continuously generated content.

Where a metaverse platform's generative tool produces a virtual item that incorporates a registered brand (having learned the pattern during the training phase), there is no act of reproduction, since no human being has deliberately sought to copy the plaintiff's mark. This, therefore, goes beyond the "use" of a registered trademark as provided under Section 10(4) of the UK Trade Marks Act, 1994, and Section 29(6) of the Indian Trade Marks Act. Both of these statutes list down the various acts which constitute the "use" of a registered trademark; however, they fail to address situations where an AI system independently "learns" patterns and subsequently produces items that bear a striking resemblance to a registered trademark. It may be argued that Section 10(4) of the UK Trade Marks Act, 1994, and its Indian counterpart, Section 29(6), were drawn with the mindset of a deliberate human act in view. However, even that is not explicitly provided for in either statute. Put another way, there is no provision in either statute that speaks of a scenario where the "user" of the mark is not a natural or artificial person.

5.1 The Cross-Border Nature of the Metaverse and the Problem of Jurisdiction

The metaverse poses a particular challenge to the policing of counterfeit goods, precisely because it erodes the territorial anchors upon which trademark and IP enforcement typically rely: a counterfeit virtual

[14] Brian M Wheeler, *Trademarks in the Age of AI: The Emerging Legal Battlefield for Brand Owners and Users of Generative AI*, ATKINSON, ANDELSON, LOYA, RUUD & ROMO, BUSINESS LAW JOURNAL (June 23, 2025), https://www.aalrr.com/Business-Law-Journal/trademarks-in-the-age-of-ai-the-emerging-legal-battlefield-for-brand-owners-and-users-of-generative-ai; *see also* Dominique E Smith & Daniel P Mullarkey, *Trademark Risks in the AI Age: Navigating Infringement, Dilution and Genericness*, POLSINELLI (Oct. 29, 2025), https://www.polsinelli.com/publications/trademark-risks-ai-age-navigating-infringement-dilution-genericness.

[15] Google LLC v. DRS Logistics (P) Ltd., 2023 SCC OnLine Del 4809.

[16] Google LLC v. MakeMyTrip (India) Private Limited, 2023 SCC OnLine Del 7965.

[17] DRS Logistics (P) Ltd. v. Google India Pvt. Ltd., CS(COMM) 1/2017, 2026:DHC:5102 (Del. HC June 15, 2026).

sneaker or handbag 'sold' on a metaverse platform may have been minted by a developer in one jurisdiction, hosted on servers in another, paid for in a wallet registered in a third, and 'worn' by an avatar accessible to the users everywhere simultaneously. Unlike the physical world, the metaverse provides no fixed location for the infringement to have occurred, because every point of access is effectively a point of sale. Consequently, the threshold question predating the emergence of the metaverse but sharpened by it is which court has jurisdiction over the infringement claim, and against whom?

Courts confronting internet based infringement have historically resolved this via a graduated inquiry into the 'interactivity' of the online conduct: the foundational articulation is the sliding-scale test from Zippo Manufacturing Co. v. Zippo Dot Com, Inc.[18], which allocated websites onto a spectrum of websites, ranging from 'passive' websites that simply provide information and support no personal jurisdiction, through those which are 'interactive' in the middle, where jurisdiction turns on the level of exchange of commercial information, to sites where defendant knowingly and repeatedly transmits files or carries out transactions with forum residents. A metaverse storefront which allows avatars to browse, customise, and purchase counterfeit virtual goods sits squarely at the interactive-to-active end of the spectrum, making the Zippo framework a natural starting point, but exposing its limits, since it was designed for a different context of static websites.

A second, complementary line of authority is the 'effects test' from Calder v. Jones[19], where the US Supreme Court held that a forum may exercise personal jurisdiction over a non-resident defendant where the defendant's conduct is expressly aimed at the forum and causes harm that the defendant knew would be suffered there. Applied to the metaverse, this recasts the inquiry from the mechanics of the platform to the foreseeable locus of injury: if a counterfeit selling avatar or platform knowingly targets consumers or a brand owner's reputation in a particular jurisdiction, then the effects test allows that forum to assert authority, even without a physical or server based presence there.

Indian courts have had to reconcile both approaches, and the leading authority is Banyan Tree Holding (P) Ltd. v. A. Murali Krishna Reddy[20]. There, the Delhi High Court held that mere accessibility of an allegedly infringing website within the forum (Delhi) is not sufficient to establish jurisdiction under Section 20 of the CPC in a passing off or trademark action: the plaintiff must show that the defendant 'purposefully availed' itself of the jurisdiction of the forum and that this targeting caused actual injury, either commercial or reputational, within that forum. The Court expressly warned against uncritical importation of the Zippo scale, noting that interactivity alone, without demonstrable targeting and effect, cannot found jurisdiction, particularly at the pre-trial injunction stage, where the plaintiff bears the burden of a prima facie showing of both purposeful availment and resultant harm.

Read together, these three authorities frame the jurisdictional problem the paper must confront: the metaverse is, by definition, the most 'interactive' and 'targeting capable' environment yet devised, satisfying Zippo's active end and Calder's foreseeability prong comfortably. The avatar economy is constructed purely on real-time, individualised commercial exchange. Yet Banyan Tree's insistence on purposeful availment and demonstrable in-forum injury demands more than platform-wide interactivity: it is asking for evidence that a particular counterfeiting actor targeted a particular market. This is exactly where enforcement stalls in practice: metaverse platforms are architecturally borderless but IP remedies

[18] Zippo Mfg. Co. v. Zippo Dot Com, Inc., 952 F. Supp. 1119 (W.D. Pa. 1997).
[19] Calder v. Jones, 465 U.S. 783 (1984).
[20] Banyan Tree Holding (P) Ltd. v. A. Murali Krishna Reddy, CS(OS) 894/2008 (Del. HC 2009).

are doggedly territorial, forcing brand owners to a fragmented, jurisdiction-by-jurisdiction pursuit of infringers who don't reside in any jurisdiction at all.

5.2 Similarity and Confusion at Machine Scale

In Cadila Health Care Ltd v Cadila Pharmaceuticals[21] the Supreme Court outlined seven factors for assessing deceptive similarity, which include the nature of the marks, the degree of resemblance, the nature of the goods, the similarity of the goods' character and purpose, the class of purchasers, the mode of purchasing and other surrounding circumstances. The test is based on a comparison between two comparably fixed marks and the Court's own emphasis on protecting consumers from confusion which is heightened in that case by the public health stakes of pharmaceutical mislabelling, is reflective of a paradigm of one product measured against one mark.

A generative model can produce thousands of subtly distinct renderings of a branded good from a single prompt, none identical to any prior output and none identical to the registered mark, yet all individually capable of evoking the brand's trade dress. Applying Cadila's factor-by-factor comparison to a single generated instance may yield a finding of no deceptive similarity even where the pattern of generation, taken as a whole, systematically trades on a brand's goodwill. The seven factor test was not built to assess a generative pattern; it was built to assess a discrete pair of marks.

Getty Images v Stability AI is instructive here as well. The UK High Court's finding of infringement was expressly described as "historic and extremely limited," confined to particular instances where the model reproduced Getty's watermark in a manner capable of causing confusion. The judgment did not, and on the pleadings before it could not, address the systemic pattern of generation. This is precisely the limitation that a Cadila style, instance-by-instance test would replicate if applied to Indian virtual counterfeiting disputes.

A further Cadila factor, the nature and similarity of the goods themselves, assumes that both the registered mark and the impugned use attach to comparable categories of goods or services. A generated avatar skin or virtual handbag is not, in any conventional sense, in the same category of goods as the physical product the registered mark denotes. It is a rendering, often free or nearly costless to produce, with no supply chain, no manufacturing cost and no physical scarcity. Indian courts have not yet had occasion to decide whether a virtual asset is a "similar good" to its physical counterpart for infringement purposes, and the Nike v StockX litigation's unresolved "digital good versus digital receipt" debate shows that even well resourced foreign litigants have not settled the point.

5.3 Dilution Without Confusion

Because generated virtual assets frequently fail the "similar goods" threshold, a confusion based cause of action under Section 29(1) or (2) may simply not be available. However, Section 29(4) of the Trade Marks Act protects marks with an established reputation in India against use on dissimilar goods or services where that use takes unfair advantage of, or is detrimental to, the distinctive character or repute of the mark, without requiring proof of consumer confusion at all. On its face, this is a better doctrinal fit for virtual counterfeiting. A luxury handbag brand's reputation can be diluted by a flood of near identical AI-

[21] Cadila Health Care Ltd. v. Cadila Pharm. Ltd., (2001) 5 SCC 73.

generated virtual renderings even if no metaverse user is actually confused into thinking the platform is the brand's authorised seller.

In Rolex SA v Alex Jewellery Pvt Ltd[22] the Delhi High Court granted interlocutory relief under Section 29(4) of the Trade Marks Act, holding that a mark's established reputation in India evidenced through long use, extensive advertising and international registration sufficed to trigger dilution protection even outside the mark's registered goods category, without any need for the mark to first be listed as "well-known" by the Registrar under Section 11. This built on the Supreme Court's earlier recognition, in N.R. Dongre v Whirlpool Corporation[23], of transborder reputation as a basis for protecting a foreign mark even before extensive use within India. Together, these cases show Indian courts willing to protect brand distinctiveness across dissimilar goods categories without insisting on confusion. This can be considered as a doctrinal posture well suited, in principle, to virtual assets that resemble a brand's trade dress without competing in the same market.

Dilution doctrine still requires a court to identify the impugned "use" and to assess "unfair advantage" or "detriment" against a specific instance of that use. It therefore inherits the instance-by-instance limitation. A court can find dilution in one generated output, or a documented pattern of outputs put before it as evidence, but has no existing mechanism to address the platform's generative capacity as such, independent of specific instances a plaintiff happens to capture and plead. Reframing the cause of action from confusion to dilution therefore removes one obstacle, the similar goods requirement, without removing the deeper one, the need for a discrete, provable act of use.

5.4 Design Law as a Parallel and Also Insufficient Regime

A brand's protected trade dress is often, or sometimes alternatively, protectable as a registered design. Section 2(d) of the Designs Act, 2000 protects the features of shape, configuration, pattern, ornament, or composition of lines or colours applied to an article, judged solely by the eye but expressly excludes anything that is itself a trademark. In principle, a metaverse platform's generative reproduction of a handbag's distinctive silhouette could trigger a similar protection as a registered design as readily as it could trade dress under the Trade Marks Act, giving a rights holder a second statutory route into the same underlying problem.

Indian courts have held that design and trademark protection for the same feature are not mutually exclusive. In Mohan Lal v Sona Paint & Hardware[24], a Full Bench of the Delhi High Court held that registering a feature as a design places no limitation on its use as a trademark by the registrant, so that a passing off action lies once that use has generated goodwill. Composite suits for design infringement and passing off for the same trade dress are not uncommon, as the Delhi High Court's own jurisprudence in Crocs Inc USA v Bata India Ltd & Ors[25] shows, but they are limited to elements beyond the registered design. Design registration also has a scale problem. Infringement under the Designs Act also requires a comparison between the infringing article and the registered design, which involves the same sort of discrete comparison issue that the Cadila and Louboutin tests grapple with. Adding a third statutory

[22] Rolex SA v. Alex Jewellery Pvt. Ltd., 2009 (41) PTC 284 (Del).
[23] N.R. Dongre v. Whirlpool Corp., AIR 1995 Del 300, 1995 (34) DRJ 109, *aff'd*, (1996) 5 SCC 714.
[24] Mohan Lal v. Sona Paint & Hardware, AIR 2013 Delhi 143.
[25] Crocs Inc. USA v. Bata India Ltd., 2019 (78) PTC 1 (Del) (decided Jan. 18/24, 2019).

regime only adds a third forum for litigation without resolving the underlying issue that none were designed to handle continuously generated articles rather than discretely manufactured ones.

6 PLATFORM ACCOUNTABILITY UNDER SECTION 79

6.1 The Safe Harbour and the Active-Participant Test

Section 79 of the Information Technology Act, 2000, which provides a safe harbour for intermediaries, was also engaged in Christian Louboutin SAS v Nakul Bajaj & Ors[26]. The Delhi High Court found that an e-commerce portal was not insulated by Section 79's safe harbour when it went beyond a mere directory function by guaranteeing authenticity, utilising the plaintiff's trademarks in its meta tags, or exercising control over fulfilment, all of which constituted "active participation" in the infringement. The Court provided an extensive but non-exhaustive list of activities that could be indicative of such participation.

A platform that makes use of its own generative tool to produce an infringing asset in response to a user's query arguably goes beyond even the conduct outlined in the Louboutin test. The platform does not merely facilitate a transaction between a buyer and seller, it actively creates the infringing content through its own tool, which is then profited from by the platform. This appears to place it even further up the slope towards "active participation" than even a curating marketplace would, since the tool itself, not an independent seller, is the source of the infringement.

The Louboutin test was framed in terms of actions that a company could take to facilitate or benefit from infringement, each of which involved some exercise of discretion on the part of the curator. Generative AI infringement, on the other hand, arises from a model's statistical propensities, which may not be the result of any particular intention on the part of the curator. Courts evaluating a Section 79 claim in the context of a generative AI dispute would have to determine whether the mere operation of a system that regularly produces infringing content constitutes "active participation" in infringement, which the Louboutin jurisprudence did not directly address.

Section 79(3)(b) of the IT Act, which governs the safe harbour, mandates that an intermediary must expeditiously remove unlawful content upon receiving actual knowledge or awareness of its existence. In Shreya Singhal v Union of India[27], the Supreme Court directed that the "expeditious" removal of unlawful content would normally entail taking action upon receipt of a court order or a government directive; a private complaint would not be sufficient to satisfy the requirements of Section 79(3)(b). This places the onus of detecting unlawful content on the complainant rather than the intermediary, which has to remove it only once it receives such an order or directive. The notice and takedown regime which forms the basis of Section 79 does not seem to apply as directly to a generative programme as it does to a human moderator. If a trademark owner were to notify a generative programme operator that a certain output infringes upon their mark, the operator would be entitled to remove that particular output, but the mere removal of that output would not necessarily be sufficient to prevent future infringing outputs, since those would constitute separately infringing acts by the operator. The next infringing output could be completely different from the first one, and the trademark owner's knowledge of the first output would not be sufficient for them to be entitled to remove the second one. In short, the notice and takedown

[26] Christian Louboutin SAS v. Nakul Bajaj, 2018 SCC OnLine Del 12215.
[27] Shreya Singhal v. Union of India, (2015) 5 SCC 1.

regime leaves trademark owners in a weaker position than they would like; they cannot simply rely on Section 79 to prevent a continuous stream of infringing outputs so long as those are not specifically notified to the programme operator.

6.2 India's Emerging Regulatory Response to Generative AI

India has been quick to respond to the phenomenon of generative AI, even if not specifically in the context of IP, in its own way. The Ministry of Electronics and Information Technology issued advisory guidelines on 1 March 2024 for intermediaries and platforms that use AI, generative AI, or algorithms. These guidelines stated that such entities must ensure that their computer resources do not permit any bias or discrimination or threaten the integrity of the electoral process, and that under-tested or unreliable AI models be made available only with the explicit permission of the Government and after labelling the possible fallibility or unreliability of their output. Further, the advisories indicated that such platforms must label, or embed with permanent unique metadata or an identifier, synthetically created content that could be used as misinformation or a deepfake. Industry concern led to a revised advisory on 15 March 2024, which replaced the obligation to obtain prior government permission with a labelling requirement, although it remained anchored to Rule 3 of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. Both built on an earlier advisory of 26 December 2023 on misinformation and deepfakes.[28] All in all, these advisories make it clear that MeitY views generative AI deployment as involving enhanced due diligence obligations under the current Intermediary Guidelines but have so far focused on issues related to electoral integrity, bias, and unlawful content, leaving open the regulatory space for trademark and trade-dress reproduction.

7 Comparative Perspectives

7.1 European Union

In the joined references Louboutin v Amazon[29], the CJEU ruled that an online marketplace may be held liable for trademark infringement by third-party sellers if an average, reasonably informed consumer may believe that the marketplace itself is the source of the goods on its site, such as where the marketplace prominently displays its own trademarks alongside listings and offers its own delivery services for the goods. A similar "average consumer" test appears in the EU's Digital Services Act (Regulation 2022/2065)[30], which in Art 6(3) withdraws the hosting exemption for consumer protection liability where a platform's presentation "would lead an average consumer to believe that the information, or the product or service that is the object of the transaction, is provided either by the online platform itself or by a recipient of the service who is acting under its authority or control." This consumer perception based test is potentially more directly applicable to the liability of Gen AI platforms than India's "conduct based

[28] Ministry of Electronics and Information Technology, Advisory No. eNo.2(4)/2023-CyberLaws–3 (Mar. 1, 2024), *superseded by* Advisory No. eNo.2(4)/2023-CyberLaws–3 (Mar. 15, 2024); Ministry of Electronics and Information Technology, Advisory No. eNo.2(4)/2023-CyberLaws–2 (Dec. 26, 2023); Press Information Bureau, *MeitY Issues Advisory to All Intermediaries to Comply with Existing IT Rules* (Dec. 26, 2023), https://pib.gov.in/PressReleseDetailm.aspx?PRID=1990542.

[29] Cases C-148/21 & C-184/21, Louboutin v. Amazon (CJEU Dec. 22, 2022).

[30] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277) 1.

active participant" test, as it focuses on the impression created by the former's presentation rather than actions.

7.2 United States

The MetaBirkins case illustrates the US courts' approach to trademark infringement of expressive digital works, which is governed by the Rogers test[31], under which artistic use of a trademark is permissible if it has artistic relevance and is not explicitly deceptive. Nike v StockX, which was settled without a ruling on the NFT specific claims, raises more foundational issues about whether an AI or blockchain mediated representation of a branded good constitutes a "good" for trademark law purposes, and if so, what constitutes "use" of the mark.

7.3 United Kingdom

Getty Images v Stability AI[32] is the closest analogy to the situation raised by Nike and MetaBirkins in UK law. The narrow finding of infringement in Getty Images v Stability AI highlights the limited ways in which current trademark law can address AI training data and output, focusing as it did on the reproduction of watermark content, a fortuitous fact particular to the case. The UK courts have yet to grapple directly with the issue of trade dress reproduction by generative AI.

7.4 Singapore: A Soft-Law Comparator

A fifth jurisdiction merits brief discussion as it takes a significantly different approach to the issue: Singapore. Unlike the US, EU, UK, and India, which have grappled with the issue in courts, Singapore has thus far responded to the threats to IP posed by generative AI at the policy level. In particular, the Infocomm Media Development Authority and the AI Verify Foundation have issued a Model AI Governance Framework for Generative AI on 30 May 2024[33], after a public consultation begun in January 2024.

The framework contains nine pillars for trustworthy AI, among which are intellectual property and content provenance risk management measures, including recommendations for content watermarking and provenance tracking rather than liability defaulting. Although Singapore's courts have not yet addressed the issue, its approach of developing a voluntary, advisory framework for content provenance risk-management for generative AI offers useful insight nonetheless. It demonstrates that a soft law approach, of the type that India's Ministry of Electronics and Information Technology has begun to develop for other AI risks, may serve as an effective starting point for addressing trademark and content provenance concerns without waiting for court cases to intervene.

None of these jurisdictions have developed a comprehensive legal framework for holding platforms liable for their users' AI driven prompt triggered counterfeits. India is further differentiated within this group by the lack of coherence between TM and IT law, which has complicated enforcement in both physical and digital goods spaces. The EU's average consumer test, the US's Rogers test, and the UK's narrow infringement finding all address some aspects of the liability risk, but none propose a unified solution that incorporates a threshold "use" test, similarity/dilution test, and platform accountability test.

---

[31] Rogers v. Grimaldi, 875 F.2d 994 (2d Cir. 1989).

[32] Getty Images (US) Inc. v. Stability AI Ltd., [2025] EWHC 2863 (Ch).

[33] Infocomm Media Development Authority, *Artificial Intelligence in Singapore*, IMDA, https://www.imda.gov.sg/about-imda/emerging-technologies-and-research/artificial-intelligence (last visited Sept. 23, 2026).

8 TOWARD A COORDINATED ENFORCEMENT FRAMEWORK

8.1 Clarifying "Use" for Generative Output

Section 29 could be clarified to bring it in line with the "invisible keyword" rulings in DRS Logistics and MakeMyTrip, which held that a search engine's use of a trademarked keyword to improve search results, even when the keyword was not visible to the user, constituted "use" of the mark. Similarly, a platform's generative reproduction of a registered trademark, including trade dress and getup, should be considered "use" regardless of whether the mark was explicitly named in the user prompt. Such a clarification would be consistent with the generative AI specific amendments made to the IT Rules in 2026, which explicitly require notice and takedown of AI-generated deepfake audiovisual content. It should also extend to include trade dress reproduction, as suggested by the prompt without name problem, and should apply to Section 29(1), 29(2) and 29(4) dilution claims to avoid the similar goods exception.

8.2 A Generative-Active-Participant Test under Section 79

The Louboutin style factors could be supplemented with a presumption of reduced safe harbour for platforms whose users generate infringing content using the platforms' own tools, as opposed to independently uploaded content, thus recognising the former as the active participants in infringement. This presumption should be rebuttable by evidence of the platforms' documented filtering safeguards, thus distinguishing between platforms that take reasonable measures to prevent infringement and those that do not. The generative active participant presumption would dovetail with the clarified "use" test to establish direct liability for the platforms that host or distribute infringing AI content. It would also address the knowledge standard by shifting the burden of proof from whether the platform knew of a particular infringing output to whether it took reasonable steps to prevent such outputs.

8.3 Institutional Coordination

A referral mechanism between the Trade Marks Registry, the Central Consumer Protection Authority, and the Ministry of Electronics and Information Technology would allow trademark infringement complaints to be processed once, rather than split between different agencies with differing jurisdictions. This would reduce the burden on trademark owners, who would not have to pursue parallel remedies under the TM Act and the IT Act. In effect, it would allow such complaints to be directed to a single nodal agency.

8.4 Extending MeitY's Due-Diligence Advisories to Brand Protection

Rather than develop a new legal framework, India could look to extend the due diligence advisories already issued by MeitY for other areas of AI governance to trademark and trade dress protection. Metaverse platforms that employ generative AI tools could be required to adopt brand- and get-up-specific filtering measures comparable to the prompt filtering guidelines already adopted voluntarily by some AI developers to avoid deepfake harms. This would utilise an existing advisory mechanism to establish a minimum standard of liability avoidance for platforms hosting AI generated content, without requiring new legislation. It would also serve as a basis for further clarifications on trademark and trade dress harms

that MeitY may seek to issue separately. This recommendation is potentially the most immediately viable, as MeitY has already issued two advisories concerning generative AI in 2024.

While India's IT (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026[34], represent a major step towards regulating synthetic media and require notice and takedown of AI-generated audiovisual content, the rules are primarily aimed at deepfakes, impersonation, and misinformation, with limited guidance on AI driven trademark infringement. Rule 3(1)(d) and 3(3)(a)(ii) of the amended rules apply primarily to AI generated deepfake content and require traceability, embedded provenance metadata, and a compressed takedown window for such content.

8.5 Verification and Audit

This suggests a system of periodic third-party audits of platform filtering mechanisms, with the ability to rely on actual content rather than self reported compliance. These could be combined with the Trade Marks Registry's own well known marks determinations under Section 11 to create a feedback loop between the Registry and the platforms.

The preceding recommendations are not mutually exclusive; rather, they suggest a potential sequence of reforms that would collectively address the enforcement challenges associated with AI counterfeiting in the metaverse. Statutory clarification of the "use" standard would establish a legal baseline for trademark infringement by AI generated content. The "generative-active-participant" presumption would supplement it by addressing the safe harbour concerns for platforms that host or distribute such content. Institutional coordination would reduce the procedural and jurisdictional barriers for trademark owners seeking remedies, while audit and verification mechanisms would provide a supervisory function over the entire system and inform future policy updates.[35]

Return to the generative wardrobe tool described earlier in this paper, where a brand-blind prompt resulted in a virtual handbag infringing on a luxury label's trade dress. Under the clarified "use" standard discussed above, the platform's output would constitute "use" of the plaintiff's mark. Under the rebuttable active participant presumption, the platform would be held liable for hosting such content absent documented trade dress filtering, which would be necessary for it to adopt if it wished to continue operating in India, as suggested in Part 8.4 above. The trademark owner, for its part, would have the option to file a consolidated complaint to the single nodal agency suggested in Part 8.3, rather than separately litigating the trademark infringement, passing off, and information technology violations. The scenario described at the beginning of this paper as a doctrinal conundrum is transformed, under these recommendations, into a concrete compliance procedure with predictable consequences for the parties involved.

[34] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E) (Feb. 10, 2026), *amending* Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E) (Feb. 25, 2021).

[35] Indranath Gupta & Lakshmi Srinivasan, *Evolving Scope of Intermediary Liability in India*, 37 INT'L REV. L. COMPUT. & TECH. 294 (2023).

9 PRACTICAL IMPLICATIONS FOR RIGHTS HOLDERS AND PLATFORMS

9.1 For Brand Owners

Until such statutory and policy clarifications as discussed above are made, brand owners are not without recourse. Trademark portfolios should include registrations of trade dress and getup in addition to word marks, to capture the likelihood of AI driven infringement. Given the prompt without brand name scenario, a trademark registration covering only a brand's word mark leaves its visual elements vulnerable to reproduction by generative tools. Second, brand owners should consider building a paper trail in support of a potential dilution claim under Section 29(4) in anticipation of infringement, as demonstrated by the Rolex v Alex Jewellery case, where the court found that the plaintiff's extensive advertising, international sales, and use of trademarks in connection with its goods were relevant to its finding of dilution.

9.2 For Platforms

Platforms that employ generative tools should consider the implications of the new safe harbour rules on their operations. Verifiable trade dress and brand name filtering practice will become a necessity for any platform that wishes to continue operating in India. This is particularly true for multi jurisdictional platforms, which would benefit from adopting the most stringent of the regional standards, such as the EU's average consumer test, to avoid the need to maintain separate compliance practices for each jurisdiction.

9.3 For Regulators

The recommendations for the Trade Marks Registry and the Ministry of Electronics and Information Technology are the most immediately tangible, as they involve no new legislation at all. The regulator could utilise the existing advisory framework to send a message to the industry that trademark and trade dress reproduction fall within the scope of the IT Act's existing obligations. This could be done as early as the next quarter, without requiring any legislative change, by issuing guidelines similar to those issued by Singapore's Infocomm Media Development Authority and AI Verify Foundation in May 2024.

10 LIMITATIONS AND SCOPE

This paper offers recommendations based on a review of doctrinal and policy responses in other jurisdictions. It does not attempt to quantify the scale of the threat to trademarks posed by prompt driven AI counterfeits in the metaverse. An empirical assessment would require the cooperation of the platforms to identify the prevalence of AI generated counterfeits in their domain, something that this author is currently attempting to quantify in her ongoing research. Finally, a word on copyright: while this paper focuses on trademark law, the same AI generated works raise complex questions for copyright law, including the authorship and originality of the output. These questions fall outside of the scope of this paper but will be addressed separately in due course.

11 CONCLUSION

Treating AI-driven virtual counterfeits as a variation of existing online counterfeits understates the shift already underway in the courts in jurisdictions such as the EU, the US, and the UK. Enforcement models

that have been adequate for static, non-generative infringing content, be it "use" in the case of trademarks, "substantial similarity" for copyrights, or the "average consumer" test for passing off, may prove insufficient in the face of prompt driven, generative AI counterfeits at the scale suggested by Nike and the MetaBirkins dispute. The fragmentation of India's trademark and information-technology law adds to the challenge, making early doctrinal and institutional coordination, rather than waiting for the equivalent of the Getty Images v Stability AI dispute to occur in India, the preferable option. The five part recommendations articulated in this paper seek to contribute to that coordination by building upon the frameworks already available in the Indian law landscape.

* * *

PART III

Markets, Corporations and the Digital Economy

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.13

CHAPTER 13

Consumer Protection Against Dark Patterns in E-Commerce: Regulating Manipulative Online Choice

PRACHI SHARMA*

The rapid expansion of electronic commerce has transformed how consumers search, compare, and purchase goods and services. Alongside convenience, however, online platforms increasingly use "dark patterns" – deceptive or manipulative interface designs that influence consumers toward choices they may not otherwise make. Examples include hidden costs, pre-selected options, disguised advertisements, difficult cancellation processes, and misleading countdown timers. These practices raise an important consumer-protection question: whether existing legal frameworks are sufficiently equipped to protect genuine consumer choice in digital marketplaces. Rather than examining consumer protection in e-commerce broadly, this study focuses specifically on the effectiveness of consumer-protection law in addressing dark patterns during online transactions. The topic is significant because traditional consumer-protection principles generally assume that consumers make decisions based on clear and accessible information. Digital interfaces can challenge this assumption by deliberately shaping the way information and choices are presented. Understanding this problem is therefore essential for determining whether legal protection should extend beyond false information to manipulation of the consumer decision-making process itself. The research will adopt a doctrinal and comparative legal approach. It will examine relevant consumer-protection legislation, regulatory guidelines, judicial decisions, and enforcement actions concerning deceptive online interface practices. Particular attention will be given to the Indian legal framework, especially the Consumer Protection Act, 2019 and regulatory measures addressing dark patterns. Selected approaches from other jurisdictions, particularly the European Union, will be comparatively analysed to identify effective regulatory mechanisms. The study will assess whether existing rules provide adequate remedies, transparency requirements, and deterrence against manipulative design practices. The research argues that effective e-commerce consumer protection requires regulation not only of misleading content but also of misleading choice architecture. Strengthening legal standards concerning dark patterns could promote more transparent digital markets while preserving legitimate innovation in online commerce.

Keywords: Dark Patterns; E-Commerce; Consumer Protection; Digital Markets; Online Manipulation

1 INTRODUCTION

With the advent of technology, India has moved since 1991 from Phase 0 (offline consumers) towards the subjective goods purchase phase, with about 150 million Indians ready to begin e-commerce shopping, depending on the income, education and occupation of their households.[1] This rapid change reflected the pattern of consumers and purchasing behaviour of society. India soon became the second largest digital market at the global level with over 62 billion UPI transactions in 2022 and 172 billion UPI transactions in 2024, growing 46% year on year.[2]

A majority of Indian consumers are influenced more by the time they spend on digital media than by traditional channels. This also contributes to the sudden rise in e-commerce tendencies. Therefore, to

* Assistant Professor at Vivekananda Institute of Professional Studies, Technical Campus, New Delhi, India.

[1] Manoj Singhal, *E-Commerce in India: A Review*, INT'L J. COMPUT. APPLICATIONS (IJCA) (2012).

[2] Shivani Anand, *Decade-wise Transformation of E-Commerce in India: A Comprehensive Analysis*, 13 JETIR 105 (2025).

boost consumption, businesses rely on a range of digital tactics aimed at expanding their customer base while pulling in more data and revenue.[3] One well-known example is growth hacking, a technique that traces back to Hotmail's early tagline "Get your free email at Hotmail" appended to outgoing messages.[4] Not every such tactic is harmless, however. A particularly troubling category has come to be known as "dark patterns" – interface designs that steer users toward certain decisions by manipulating the information presented to them. These designs work by tapping into cognitive biases, such as the fear of scarcity or the pull of the bandwagon effect, to drive up revenue and push consumers toward greater consumption.[5]

As businesses have leaned more heavily on data-driven marketing and outcome-based metrics, the use of dark patterns in online retail has grown correspondingly. Firms keep close track of figures like click-through rates, cart completions, and subscription renewals and this focus tends to push them toward tactics that chase immediate gains, sometimes at the cost of keeping customers satisfied over time. The fallout is that shoppers frequently find themselves pushed into purchases they never intended to make, locked into subscriptions they didn't mean to sign up for, or misled about what exactly they're buying. Such deception chips away at the trust that underpins the relationship between consumers and brands, making it a serious concern both ethically and from a legal standpoint.

Dark patterns take many forms in practice – drip pricing, disguised advertising, bait-and-switch tactics, artificially manufactured urgency, and several others. Under Indian law, these practices are captured within the broad definition of "unfair trade practice" found in Section 2(47) of the Consumer Protection Act, 2019.[6]

Acting under its powers under Section 18 of the Act, the Central Consumer Protection Authority (CCPA) took a concrete step on 30th November 2023, releasing the "Guidelines for Prevention and Regulation of Dark Patterns, 2023." These guidelines went further than the general statutory language by naming thirteen distinct dark patterns specifically observed in the e-commerce space: false urgency, basket sneaking, confirm shaming, forced action, subscription traps, interface interference, bait and switch, drip pricing, disguised advertisements, nagging, trick wording, SaaS billing tricks, and rogue malware.[7]

Building on this regulatory groundwork, the Department of Consumer Affairs organised a stakeholder meeting on 28th May 2025, chaired by the Union Minister for Consumer Affairs, Food & Public Distribution and New and Renewable Energy.[8] The meeting brought together representatives from major e-commerce companies, industry bodies, voluntary consumer organisations, and National Law Universities to discuss ways of curbing deceptive online practices.

This dialogue led directly to a follow-up measure: on 5th June 2025, the CCPA issued an advisory under the Consumer Protection Act, 2019, calling on e-commerce platforms to carry out self-audits aimed

---

[3] Arunesh Mathur et al., *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*, 3 PROC. ACM ON HUM.-COMPUT. INTERACTION (CSCW) 1, 1–32 (2019).

[4] Harry Brignull, *Dark Patterns: Inside the Interfaces Designed to Trick You*, HTG REPORTS (2018) (originating the term "dark patterns").

[5] Arvind Narayanan et al., *Dark Patterns: Past, Present, and Future*, 18 ACM QUEUE 67, 67–92 (2020).

[6] The Consumer Protection Act, 2019, § 2(47).

[7] Central Consumer Protection Authority, Guidelines for Prevention and Regulation of Dark Patterns, 2023, F. No. J-24/9/2022-CPU, Gazette of India (Nov. 30, 2023), issued under § 18 of the Consumer Protection Act, 2019 (Annexure 1 lists the thirteen specified dark patterns).

[8] Press Release, Ministry of Consumer Affairs, Food and Public Distribution, Government of India, *Stakeholder Meeting on Eliminating Deceptive Online Practices* (May 28, 2025), https://pib.gov.in.

at detecting dark patterns on their own sites, with the broader goal of fostering a fair, ethical, and consumer-oriented digital marketplace.[9]

Through this advisory, e-commerce platforms were directed to take active measures against engaging in dark-pattern-style deceptive practices. They were also given a three-month window from the date of the advisory to complete self-audits identifying any such patterns and to remedy them. Once these audits were done, platforms were expected to issue self-declarations confirming they were free of dark patterns, a step intended to reinforce trust between businesses and consumers while supporting a more transparent digital ecosystem.[10]

Alongside this, a Joint Working Group was formally set up through an Office Memorandum dated 5th June 2025.[11] Comprising representatives from various ministries, National Law Universities, and voluntary consumer organisations, the group's mandate is to identify dark patterns collaboratively and bring stakeholders together to build a transparent, ethical, and user-focused online environment.

Are existing consumer-protection laws in India effective enough to deal with dark patterns in e-commerce, or do digital manipulative practices require stronger, different legal regulation? This research dives into analysis of existing Indian laws and thereafter, selected approaches from other jurisdictions, particularly the European Union, will be comparatively analysed to identify effective regulatory mechanisms.

2 Analysing the Effectiveness of Existing Consumer Protection Laws in India

One study drew on a labelled dataset of over 50,000 text samples pulled from e-commerce websites, achieving 93.2% accuracy in flagging manipulative language patterns. Commonly identified deceptive phrases included false scarcity cues such as "Only a few left in stock!", confirm-shame prompts like "Are you sure you want to miss this deal?", and misdirection tactics such as "By continuing, you agree to our terms." A parallel visual analysis, trained on more than 10,000 images of checkout flows, pop-ups, and other UI elements from e-commerce platforms, achieved an 89.5% detection rate for deceptive visual features including greyed-out cancel buttons, fake countdown timers, and pre-ticked opt-in checkboxes.[12]

These high accuracy rates suggest that AI-driven detection tools offer a viable route for monitoring dark patterns at scale. Beyond their immediate effect on purchasing behaviour, dark patterns were found to erode consumer trust over time. In the short term, users tend to be pressured into transactions without realising they are being manipulated; in the long term, however, this manipulation breeds distrust, driving customers away and generating negative reviews. The study reported that 74% of surveyed users expressed concern about manipulative design tactics, while 49% said they would avoid returning to a website if they felt deceived by it and platforms that relied heavily on dark patterns tended to show weaker customer retention overall.[13] Taken together, these findings suggest that although dark patterns may yield

[9] Central Consumer Protection Authority, Advisory in terms of Consumer Protection Act, 2019 on Self-Audit by E-Commerce Platforms for Detecting the Dark Patterns on their Platforms to Create a Fair, Ethical and Consumer Centric Digital Ecosystem (June 5, 2025), https://www.pib.gov.in/PressReleasePage.aspx?PRID=2134765.
[10] *Id.*
[11] Ministry of Consumer Affairs, Food and Public Distribution, Government of India, Office Memorandum Constituting a Joint Working Group on Dark Patterns (June 5, 2025).
[12] S.V.J. Manikanta Gupta et al., *Identifying and Mitigating Dark Patterns in E-Commerce Websites* 3 (SSRN, Nov. 30, 2025), https://ssrn.com/abstract=5832222.
[13] *Id.*

short-term sales gains, they ultimately come at the cost of brand reputation and long-term customer loyalty.

The working of the CCPA is best illustrated through its enforcement record rather than through the guidelines alone. Beyond issuing advisories and guidance documents, the Authority has begun translating these standards into concrete penalties against platforms found violating them.[14] In one instance, an online coaching platform was penalised to the tune of Rs 3,00,000 for displaying a countdown timer that gave the false impression of a time-bound offer, a tactic meant to rush buyers into decisions they might not otherwise make so quickly. In another case, a baby-products retailer was fined Rs 2,00,000 for a pricing practice known as drip pricing – the platform had displayed prices as inclusive of tax, only to add GST separately at the point of checkout, leaving buyers to discover the actual cost later than they should have. A third instance involved an online pharmacy platform, penalised Rs 1,00,000 for quietly adding a paid membership to users' shopping carts without seeking their explicit approval beforehand.[15]

The clearest instance of enforcement to date remains the CCPA's Rs 7 lakh penalty imposed on Zepto Marketplace, issued through an order dated 4th December 2025 for two identified violations: drip pricing, where handling charges were disclosed only at the final stage of checkout rather than upfront, and basket sneaking, involving the automatic or pre-selected addition of a paid membership to the customer's cart. Following the order, the platform discontinued both practices.[16] Zepto has since challenged the penalty before the National Consumer Disputes Redressal Commission (NCDRC), securing an interim stay on 20th January 2026, and the matter remains pending before the tribunal. Government statements made to Parliament in August 2026 listed this penalty among those collected and confirmed that the flagged features had been removed by the company. A separate notice issued by the CCPA on 4th June 2025, concerning alleged overcharging beyond the Maximum Retail Price (MRP) and alleged disguised advertisement, is also reported to remain pending.[17]

Complaints surrounding differential pricing have received similar public attention. Consumer testing and independent reviews have pointed to instances where identical products were priced higher for iOS users than for Android users, a pattern also observed on comparable platforms. Regulatory authorities have sought clarifications on this practice, though no conclusive, platform-specific finding of violation has yet been recorded.[18]

If substantiated, complaints of this kind on a high-frequency delivery platform would point to recurring consumer harm of the sort that the law on unfair trade practices is designed to address.

These examples reflect a broader pattern in how the CCPA operates: it does not merely lay down definitions and categories of dark patterns on paper but follows through with monetary penalties once

---

[14] Himanshi Hans, *CCPA Fines Zepto, Physics Wallah, BookMyShow for Using 'Dark Patterns' to Mislead Consumers*, LAWBEAT (Aug. 7, 2026), https://lawbeat.in/news-updates/ccpa-fines-zepto-physics-wallah-bookmyshow-for-using-dark-patterns-to-mislead-consumers-1620072 (reporting on written reply by B.L. Verma, Minister of State for Consumer Affairs, Food and Public Distribution, in the Rajya Sabha).

[15] Rajya Sabha, Unstarred Question, Ministry of Consumer Affairs, Food and Public Distribution, Reply by B.L. Verma, Minister of State for Consumer Affairs, Food and Public Distribution, on CCPA Enforcement Action Against Dark Patterns (Rajya Sabha, Aug. 6, 2026).

[16] *Id.*

[17] *Zepto, BookMyShow, IndiGo Among Nine Fined Nearly Rs 20 Lakh by CCPA Over Dark Patterns*, BESTMEDIAINFO (2026), https://bestmediainfo.com/mediainfo/mediainfo-marketing/zepto-bookmyshow-indigo-among-nine-fined-nearly-rs-20-lakh-by-ccpa-over-dark-patterns-12236513.

[18] *Inside India's Dark Pattern Crackdown: Why Zepto, BookMyShow and Others Are Under CCPA Scanner*, OUTLOOK BUSINESS (2026), https://www.outlookbusiness.com/news/inside-indias-dark-pattern-crackdown-why-zepto-bookmyshow-and-others-are-under-ccpa-scanner.

violations are identified, signalling that compliance is being actively monitored rather than left to voluntary adherence alone.

3 COMPARATIVE APPROACH WITH THE EUROPEAN UNION

The first source under review is a guidance document on deceptive interface design, released in March 2022 by the European Data Protection Board (EDPB) for member states within the European Union. Its focus lies in offering developers practical direction on designing social media interfaces in a manner that avoids manipulative practices. Within this document, dark patterns are organised into six broad categories, further broken down into fifteen sub-categories, each illustrated with concrete examples. This classification framework is anchored in core data-protection principles – lawfulness, fairness, transparency, purpose limitation, and data minimisation – as they apply to interface design choices. Enforcement against violations identified under this framework operates alongside, and draws its authority from, the General Data Protection Regulation (GDPR), 2016.[19]

TABLE 1. *Regulation of dark patterns: CCPA (India) and EDPB (EU) compared*

Aspects	**CCPA (India)**	**EDPB (EU)**
Issuing Authority	Central Consumer Protection Authority, under the Consumer Protection Act, 2019	European Data Protection Board, under the GDPR, 2016
Year of Issuance	2023 (Guidelines); 2025 (Advisory)	2022 (v1.0); finalised 2023 (v2.0)
Legal Basis	Consumer Protection Act, 2019 (Section 18)	General Data Protection Regulation (GDPR), 2016
Primary Focus	Consumer protection and fair trade practices in e-commerce	Data protection and privacy in social media interfaces
Scope of Application	All e-commerce platforms	Primarily social media platforms (though broader relevance acknowledged)
Classification	13 specified dark patterns	6 categories, 15 sub-categories
Underlying Principles	Prevention of unfair trade practices, transparency, consumer trust	Lawfulness, fairness, transparency, purpose limitation, data minimisation
Compliance Mechanism	Self-audit by platforms + self-declaration	Recommendatory guidance for interface design
Enforcement Mechanism	Enforcement under Consumer Protection Act, 2019	Enforcement linked to GDPR provisions
Nature of Document	Regulatory guidelines + advisory	Guidelines (non-binding recommendations)

The table above sets out a comparative overview of the regulatory approaches adopted by India and the European Union in addressing dark patterns, highlighting the divergences and overlaps between the two frameworks. In India, the Central Consumer Protection Authority (CCPA), acting under the powers conferred by Section 18 of the Consumer Protection Act, 2019, has issued a dedicated set of guidelines identifying thirteen specific dark patterns prevalent in the e-commerce sector, later reinforced through a compliance-driven advisory mandating self-audits by platforms. This approach is rooted primarily in

[19] EUROPEAN DATA PROTECTION BOARD, GUIDELINES 03/2022 ON DECEPTIVE DESIGN PATTERNS IN SOCIAL MEDIA PLATFORM INTERFACES: HOW TO RECOGNISE AND AVOID THEM (Mar. 14, 2022) (version 1.0, for public consultation), https://www.edpb.europa.eu.

consumer-protection law, with an emphasis on preventing unfair trade practices and preserving consumer trust in digital transactions.[20]

By contrast, the European framework, represented through the European Data Protection Board's (EDPB) guidelines, approaches the issue predominantly through the lens of data protection rather than consumer protection, situating dark patterns within the broader architecture of the General Data Protection Regulation (GDPR).[21] Here, the concern is less about unfair commercial practices in the transactional sense and more about how deceptive interface design undermines core data-protection principles such as lawfulness, fairness, transparency, purpose limitation, and data minimisation. The EDPB's classification is also comparatively more granular, spanning six overarching categories and fifteen sub-categories, and its scope is centred on social media platforms rather than e-commerce broadly, even though its principles carry wider applicability.[22]

Taken together, the comparison illustrates two distinct regulatory philosophies: the Indian model treats dark patterns chiefly as a consumer-protection concern enforced through sector-specific guidelines and compliance obligations, whereas the EU model treats them as a data-protection concern enforced through the overarching machinery of the GDPR.[23] This distinction is significant for the present study, as it demonstrates that the effectiveness of legal protection against dark patterns may depend considerably on which regulatory lens – consumer protection or data protection – a jurisdiction chooses to adopt.

4 CONCLUSION AND SUGGESTIONS

Concern over regulating manipulative digital design has grown steadily, as governments and regulators grapple with new and evolving forms of online manipulation aimed at swaying user behaviour. India took a meaningful step in this direction through the Consumer Protection Act, 2019, followed by the Dark Pattern Guidelines, 2023, both of which mark a significant move toward reinforcing consumer safeguards within the digital economy. Through these measures, the government has sought to curb deceptive practices online while fostering greater transparency in how digital transactions unfold.

That said, legislation alone cannot fully root out such practices. For the law to serve its intended purpose, it must be paired with robust enforcement, heightened public awareness, and an ongoing capacity to keep pace with technological change. As digital platforms lean more heavily on artificial intelligence and increasingly personalised interfaces, regulators will need to stay alert to fresh and evolving forms of manipulation that may emerge. This piece has sought to demonstrate that while the existing legal framework represents meaningful progress, its real effectiveness hinges on stronger enforcement mechanisms, more precise compliance benchmarks, and regular review of the law to keep it responsive

[20] Central Consumer Protection Authority, Guidelines for Prevention and Regulation of Dark Patterns, 2023, Gazette of India (Nov. 30, 2023); Central Consumer Protection Authority, Advisory in terms of Consumer Protection Act, 2019 on Self-Audit by E-Commerce Platforms for Detecting the Dark Patterns on their Platforms to Create a Fair, Ethical and Consumer Centric Digital Ecosystem (June 5, 2025).

[21] EUROPEAN DATA PROTECTION BOARD, GUIDELINES 03/2022 ON DECEPTIVE DESIGN PATTERNS IN SOCIAL MEDIA PLATFORM INTERFACES: HOW TO RECOGNISE AND AVOID THEM (Feb. 14, 2023) (version 2.0).

[22] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation), 2016 O.J. (L 119) 1.

[23] EUROPEAN DATA PROTECTION BOARD, *supra* note 22.

to change. Striking the right balance – one that safeguards consumers without stifling innovation – is what will ultimately shape a digital marketplace that is fairer, more transparent, and more trustworthy.

* * *

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.14

CHAPTER 14

Dark Patterns in Indian E-Commerce: Emerging Challenges for Consumer Protection and Algorithmic Accountability

PARTH SHARMA*

Indian e-commerce market has seen a rapid growth over the past decade. As the economy is undergoing structural changes, e-commerce has emerged as one of the adaptive elements considering the shift in consumer behavior, digital payments and internet access. Ensuring fair and open markets will become important to sustain the growth of e-commerce. The rapid growth in e-commerce led to convenience for the consumers but at the same time it has also exposed consumers to hidden digital manipulations known as dark patterns. Deceptive digital interfaces manipulate user decision making and compromise individual autonomy. This is facilitated by employing manipulative strategies such as artificial urgency cues, hidden transaction fees, and forced recurring billing structures.

The legal framework to deal with dark patterns revolves around the Consumer Protection Act, 2019, which authorizes the Central Consumer Protection Authority under section 18 to prohibit unfair trade practices. This framework was reinforced when CCPA issued specific guidelines on Dark Patterns that classify these manipulative designs as legal violations.

Despite these rules, there exists a gap between written laws and rapid changes in online platform code. Significant challenges remain because traditional consent rules often fail against automated systems that manipulate on a massive scale. Developers continuously alter their interfaces to bypass oversight of the regulators which is a challenge for regulatory bodies.

This paper examines the influence of manipulative interface designs on consumer choice and the effectiveness of Consumer Protection Act, 2019 and CCPA in protection of consumers in e-commerce from dark patterns. It proposes solutions to hold platforms legally accountable for the violations.

Keywords: Dark Patterns, Consumer Protection Act 2019, Central Consumer Protection Authority, Consumer Autonomy, Digital Regulation

1 INTRODUCTION

Dark Patterns are consumer-targeted marketing techniques which are designed to influence consumers' choices in online platforms. They often take advantage of cognitive biases and encourage consumers to make decisions that they may not have intended to make. The term "dark patterns" was coined by user experience designer Harry Brignull. In present times, the concept has gained attention due to rapid expansion of e-commerce and evolving nature of consumer behaviour.

These techniques use subtle nudges to influence consumers' decisions. Unlike ordinary marketing strategies, dark patterns may force consumers towards choices that benefit the business compared to their actual preferences. Digital businesses use algorithmic tools and behavioral science to influence

* Research Scholar at National University of Study and Research in Law, Ranchi, Jharkhand, India.

consumer choices. While these practices can help businesses in profit making, at the same time they can impact consumers' free and informed choices.

The Bureau of Indian Standards (BIS) has notified a framework on 'Online Consumer Reviews — Principles and Requirements for their Collection, Moderation and Publication' on 23.11.2022 for safe-guarding and protecting consumer interest from fake and deceptive reviews in e-commerce.

The standards are voluntary and are applicable to every online platform which publishes consumer reviews. The guiding principles of the standard are integrity, accuracy, privacy, security, transparency, accessibility and responsiveness.

The growing use of dark patterns raises several consumer concerns. As e-commerce is rapidly growing in India, the use of such practices presents new challenges for protecting consumers and ensuring greater accountability from digital platforms.[1]

2 CONSTITUTIONAL MANDATE AND CONSUMER PROTECTION

Consumer protection is not expressly mentioned in the Indian Constitution. However, its constitutional basis can be traced to the Preamble, Fundamental Rights and the Directive Principles of State Policy. These provisions reflect the broader constitutional commitment to justice, equality, dignity and social welfare.

The Preamble seeks to secure social and economic justice, equality of status and opportunity, and dignity for all citizens. These objectives provide a foundation for protecting consumers from exploitation and unfair practices in the marketplace.

The Fundamental Rights also have relevance to consumer protection. Article 14 guarantees equality before the law and equal protection of the laws. Article 21 protects the right to life and personal liberty. The courts have given Article 21 a wide interpretation, which includes the right to live with dignity and in a safe and healthy environment.

The Supreme Court also recognised access to justice as a fundamental right flowing from Articles 14 and 21. The Court emphasised that justice must be effective, accessible, speedy and affordable.[2] This principle is particularly relevant to consumer protection, where an effective and timely remedy is essential for addressing consumer grievances.

The Directive Principles of State Policy further strengthen this framework. Article 38 requires the State to promote the welfare of the people and secure a social order based on justice. Article 39 also seeks to prevent exploitation and promote the interests of citizens.

Therefore, although consumer protection is not expressly provided as a constitutional right, its principles flow from the Constitution and find statutory recognition through the Consumer Protection Act, 2019.

---

[1] Sandeep J. Sharma & Ishita Sharma, *Dark Patterns in a Bright World: An Analysis of the Indian Consumer Legal Architecture*, 11 INT'L J. ON CONSUMER L. & PRAC. art. 7 (2023), https://repository.nls.ac.in/cgi/viewcontent.cgi?article=1122&context=ijclp (last visited Sept. 4, 2026).

[2] Anita Kushwaha v. Pushap Sudan, (2016) 8 SCC 509.

3 UNDERSTANDING DARK PATTERNS IN E-COMMERCE

Generally, e-commerce entities, platforms and sellers resort to deceptive practices such as dark patterns to boost sales, giving priority to profits over consumer autonomy. The sections of digital consumers who are prone to these practices are minors, the elderly and those with low literacy levels.

Dark Patterns are regarded as unfair trade practices and misleading advertisements which are prohibited under the Consumer Protection Act, 2019. Though the Consumer Protection Act, 2019, E-Commerce Rules, 2020 and the Guidelines for Prevention of Misleading Advertisements address the issue of Dark Patterns, comprehensive regulations were required to tackle the modern challenges associated with it. To address this concern, the Government of India issued the Guidelines for Prevention and Regulation of Dark Patterns, 2023 under Section 18 of Consumer Protection Act, 2019. Any contravention to the guidelines will lead to penalty under the Consumer Protection Act, 2019. In case of any dispute regarding the interpretation of guidelines, the decision of the Central Consumer Protection Authority (CCPA) will be final. The guidelines are applicable to all platforms, sellers and advertisers offering goods or services in Indian markets.

4 STRATEGIES AND GUIDELINES FOR DARK PATTERNS PREVENTION

The Guidelines for Prevention and Regulation of Dark Patterns, 2023 outline practices falling under the category of dark patterns in Annexure 1. There are several practices which fall under the category of dark patterns such as false urgency, basket sneaking, confirm shaming, forced action, subscription trap, interface interference, bait and switch, drip pricing, disguised advertisement, nagging, trick question, SaaS billing and rogue malwares which can affect consumers' autonomy while making an informed decision.

A brief description of the dark patterns is mentioned below:

TABLE 1. *Dark patterns specified in Annexure 1 to the Guidelines for Prevention and Regulation of Dark Patterns, 2023*

Dark Pattern	**Brief Description**
False Urgency	Creates artificial time pressure to influence user decisions.
Basket Sneaking	Adds unwanted products or services to user's basket.
Confirm Shaming	Uses guilt or shame to discourage user refusal.
Forced Action	Requires users to perform unwanted actions to proceed.
Subscription Trap	Makes subscriptions easy to start, difficult to cancel.
Interface Interference	Manipulates interface elements to prioritize unintended user choices.
Bait and Switch	Advertises one option but delivers another instead.
Drip Pricing	Reveals additional costs gradually during the purchasing process.
Disguised Advertisement	Presents advertisements as ordinary content or genuine recommendations.
Nagging	Repeatedly prompts users to perform unwanted actions.
Trick Question	Uses confusing wording to obtain unintended user consent.
SaaS Billing	Uses confusing billing practices to obtain recurring payments.
Rogue Malware	Uses malicious software to deceive, manipulate, or harm.

The CCPA is responsible for enforcement of Guidelines on Prevention of Dark Patterns and its interpretation. The procedure to enforce the guidelines is followed as enshrined under the Consumer Protection Act, 2019 for addressing unfair trade practices. CCPA may initiate a suo motu inquiry or act

upon a consumer complaint in the event of violation of guidelines. A preliminary inquiry is conducted to ascertain if a prima facie case exists and upon confirmation, the matter is sent for investigation. If CCPA confirms violation of guidelines, it can issue an order to modify or cease the infringing practice and impose penalties. Before making a final decision, the infringing party is given an opportunity to be heard.[3]

If proved, the punishment for the offence of misleading advertisement under the Consumer Protection Act, 2019 will be applicable. The imprisonment as a punishment may extend to a term of two years and fine of up to rupees ten lakhs. In case of repeated offence, the punishment may extend to a term of five years and fine of up to rupees fifty lakhs.[4]

Although the CCPA has been playing a proactive role by taking suo motu cognizance of consumer protection issues and initiating action against unfair practices, its effective functioning continues to face several challenges. These include inadequate staffing, difficulties in penalty collection, and concerns relating to transparency and public accountability.[5]

5 CCPA ACTION AGAINST DARK PATTERNS IN DIGITAL PLATFORMS

CCPA has acted against PhysicsWallah Limited and McAfee Software India Private Limited for what it found to be deceptive interfaces that misled consumers and influenced their choices on digital platforms. PhysicsWallah was imposed a penalty of rupees 5 lakhs while McAfee was fined 1 lakh rupees. CCPA also directed both the companies to remove manipulative digital interfaces to ensure autonomy in consumers' informed choices. The action was taken by CCPA in accordance with the Consumer Protection Act, 2019, E-Commerce Rules, 2020 and the Guidelines for Prevention and Regulation of Dark Patterns, 2023.

In the case of PhysicsWallah, CCPA found that a certain amount of donation to PhysicsWallah Foundation was automatically selected during checkout and it was added to the total amount payable without consumers' explicit consent. Emotional messages relating to children's education and healthcare were also used to encourage consumers to give donations. Also, courses which were advertised free could only be accessed after users submit personal information such as contact number and email addresses. CCPA considered these as dark patterns and identified them as basket sneaking, confirm shaming and forced action.

CCPA also examined McAfee's subscription renewal process and found that consumers have not been given a neutral choice regarding renewal. The platform and interface prominently displayed the options as "renew now" or "accept risk". These kinds of options represent that non-renewal is a risky decision. According to CCPA, the phrase "accept risk" reflects that consumers could be exposed to cyber security tricks if they did not renew the subscription, although this claim cannot be guaranteed by the company. CCPA identifies this practice as confirm shaming, interface interference, forced action and

[3] Neha, *Dark Patterns Outlawed: Online Platforms to Watch Out*, SCC ONLINE BLOG (Feb. 27, 2025), https://www.scconline.com/blog/post/2025/02/27/dark-patterns-outlawed-online-platforms-to-watch-out/.

[4] Consumer Protection Act, 2019, § 89.

[5] Gaurav Pathak, Mohit Yadav & Anush Ganesh, *Regulatory Innovation Through Authority: An Analysis of India's Central Consumer Protection Authority's Functioning (2020–2025)*, 13 INT'L J. ON CONSUMER L. & PRAC. art. 4 (2026), https://repository.nls.ac.in/ijclp/vol13/iss1/4 (last visited Sept. 6, 2026).

trick question. A consumer should be given autonomy to decide without fear-based messaging or misleading design.[6]

To strengthen compliance, CCPA issued an advisory on June 5th 2025 asking e-commerce companies and digital platforms to conduct self-audits and remove dark patterns from their interfaces.

A total of 26 leading e-commerce platforms have voluntarily submitted their self-declaration letters confirming compliance with the Guidelines for Prevention and Regulation of Dark Patterns, 2023.

The platforms that have submitted declarations include Zepto, Flipkart, Myntra, MakeMyTrip, Reliance Digital, Swiggy, Zomato etc.[7]

6 INSTITUTIONAL FRAMEWORK FOR CONSUMER PROTECTION AND REDRESSAL

The Consumer Protection Act, 2019 provides for a three-tier quasi-judicial machinery at District, State and National level known as "Consumer Commissions" for protection of the rights of consumers. It aims to provide simple and speedy redressal of consumer disputes including those related to unfair trade practices.

As per Section 38(7) of the Consumer Protection Act, 2019, every complaint shall be disposed of as expeditiously as possible, and an endeavour shall be made to decide the complaint within the specified period. Where the complaint does not require analysis or testing of commodities, it should be decided within three months from the date of receipt of notice by the opposite party. Where the complaint requires analysis or testing of commodities, it should be decided within five months from the date of receipt of such notice.

To ensure speedy justice for consumers, the Consumer Protection Act provides that no adjournment shall ordinarily be granted by the Consumer Commissions unless sufficient cause is shown. The reasons for granting such adjournment must also be recorded in writing by the Commission.

The National Consumer Helpline (NCH) administered by the Department of Consumer Affairs has emerged as a single point of access to consumers across the country for their grievance redressal at a pre-litigation stage. Consumers can register their grievances from all over the country in 17 languages including Hindi, English, Kashmiri, Punjabi, Nepali, Gujarati, Marathi, Kannada, Telugu, Tamil, Malayalam, Maithili, Santhali, Bengali, Odia, Assamese and Manipuri through a toll-free number 1915.[8]

These grievances can be registered on Integrated Grievance Redressal Mechanism (INGRAM), an omnichannel IT enabled central portal, through various channels – WhatsApp, SMS, email, the NCH app, the web portal (consumerhelpline.gov.in) and the Umang app, as per their convenience. 1,853 companies, who have voluntarily partnered with NCH, as part of the 'Convergence' programme directly respond to these grievances according to their redressal process and revert by providing feedback to the complainant on the portal.[9]

---

[6] Ministry of Consumer Affairs, Food & Public Distribution, *CCPA Acts Against Dark Patterns on Digital Platforms*, PRESS INFORMATION BUREAU, GOVERNMENT OF INDIA (June 3, 2026), https://www.pib.gov.in/PressReleasePage.aspx?PRID=2268302&reg=3&lang=1.

[7] Government of India, Ministry of Consumer Affairs, Food and Public Distribution, Department of Consumer Affairs, Lok Sabha Unstarred Question No. 1628 (to be answered on Dec. 10, 2025), https://sansad.in/ls/questions/questions-and-answers (last visited Sept. 5, 2026).

[8] Government of India, Ministry of Consumer Affairs, Food and Public Distribution, Department of Consumer Affairs, Rajya Sabha Unstarred Question No. 1798, Regulation of Dark Patterns on Digital Platforms (to be answered on Aug. 4, 2026), https://sansad.in/rs/questions/questions-and-answers (last visited Sept. 4, 2026).

[9] *Id.*

7 STRENGTHENING CONSUMER PROTECTION THROUGH E-COMMERCE REGULATIONS

The Government has amended the Consumer Protection (E-Commerce) Rules, 2020, to strengthen checks on dark patterns, fake sponsored listings and price manipulation on online platforms. The Consumer Protection (E-Commerce) (Amendment) Rules, 2026, will come into force from 1 January 2027. According to the Department of Consumer Affairs, the amendment seeks to strengthen consumer protection while maintaining a transparent and balanced regulatory framework. It will also consider Ease of Doing Business and avoiding unnecessary regulatory burdens on e-commerce entities.

A key provision requires every e-commerce entity to become a partner in the convergence process of the National Consumer Helpline (NCH), thereby strengthening the connection between e-commerce platforms and the national grievance redressal mechanism. The Department stated that the NCH received 17,71,622 grievances in 2025, of which 5,11,196, or around 29 per cent, related to the e-commerce sector. The amended rules also require e-commerce entities to provide complainants with a copy of the complaint as recorded by their grievance officer.[10]

The amendments introduce requirements concerning search results, sponsored listings and price reductions. E-commerce entities cannot manipulate search results in a manner that misleads users or affects the relevance of results to their search queries. Sponsored listings must be identified through clear and prominent disclosures. Where a price reduction is announced, both the reduced price and the prior price must be displayed. The prior price is defined as the lowest price at which the goods or services were offered during the 30 days preceding the announcement.

With regard to dark patterns, e-commerce entities are required to comply with the Guidelines for Prevention and Regulation of Dark Patterns, 2023. They must also undertake a yearly self-audit and prominently display a certificate of compliance. Marketplace e-commerce entities are further required to provide important seller and product information, including best-before or use-before dates and details relating to returns, refunds, warranty, delivery and payment, to enable informed consumer decisions. They are prohibited from using consumer information for specified purposes without express and affirmative consent and from collecting bundled fees for services unrelated to the e-commerce platform, subject to the specified exception for loyalty or membership programmes. In the case of imported goods, the importer's details and country of origin must also be disclosed.

The Consumer Protection (E-Commerce) Rules, 2020, notified under the Consumer Protection Act, 2019, provide the existing framework for protecting consumers against unfair trade practices in the e-commerce sector. The Department stated that the amendments further strengthen this framework in response to evolving business models, digital practices and consumer expectations. They are intended to promote a more transparent, accountable and consumer-centric e-commerce ecosystem, while providing e-commerce entities with greater clarity regarding their responsibilities and supporting a level playing field in the digital marketplace.

[10] *Government Amends E-Commerce Rules to Check Price Manipulation, Dark Patterns, Fake Sponsored Listings*, ECON. TIMES (Sept. 10, 2026), https://economictimes.indiatimes.com/industry/services/retail/government-amends-e-commerce-rules-to-check-price-manipulation-dark-patterns-fake-sponsored-listings/articleshow/134009564.cms.

8 Conclusion

The guidelines for prevention of dark patterns play a significant role in promoting consumer rights in the digital age. Strict enforcement is crucial for protecting consumers from malpractices in e-commerce. Keeping in pace with the international best practices and also, being the forerunners of consumer rights, the Department of Consumer Affairs has acted proactively to protect consumers from such deceptive patterns. The Central Consumer Protection Authority (CCPA) in India is empowered under the Consumer Protection Act, 2019 to take various actions to ensure consumer rights are protected.

The CCPA has taken a proactive approach in addressing dark patterns, which is commendable, but its role can be made more effective through stronger monitoring, better enforcement, greater transparency and improved investigative capacity.

Apart from establishing a publicly accessible case management system and strengthening investigative infrastructure, the CCPA could take extra measures to improve its ability to address dark patterns. Regular monitoring of e-commerce websites, applications and other digital platforms could help identify manipulative practices at an early stage. The existing requirement of annual self-audits by e-commerce entities could also be strengthened through clearer reporting and compliance requirements. As digital practices continue to evolve, the CCPA could regularly review existing categories of dark patterns and provide guidance on emerging forms of manipulation.

A simple and accessible mechanism for consumers to report suspected dark patterns could further assist in identifying problematic practices. Greater accountability for repeated violations and timely investigation and disposal of cases would also strengthen enforcement. Better coordination between the CCPA, the National Consumer Helpline and e-commerce entities could help identify recurring consumer complaints and patterns of manipulation. Consumer awareness regarding common dark patterns such as Basket Sneaking, Confirm Shaming, Forced Action and Interface Interference should also be strengthened so that consumers can recognise and report such practices.

Also, regular publication of information regarding investigations, violations and compliance with CCPA directions could improve transparency and public accountability. Particular attention may also be given to digital practices affecting consumers who may be more susceptible to manipulative designs, especially students and minors.

* * *

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.15

CHAPTER 15

The Invisible Employer: Algorithmic Management and the Reinvention of Control in the Gig Economy

CHANDRALEKHA

Every employment relationship in labour law has, until now, presupposed a visible employer — a person or entity capable of being identified, questioned, and held to account for the direction it exercises over a worker. The gig economy disturbs this premise. Platforms exercise the same functions traditionally associated with an employer — assigning work, setting pay, monitoring performance, and terminating access — not through a visible supervisor but through an algorithm whose rules the worker cannot see and cannot contest. This paper asks whether the traditional control test, built on the assumption of an identifiable, human directing mind, can meaningfully regulate a relationship in which the employer's authority has become functionally invisible, and if not, how the law should respond.

The paper adopts a doctrinal and comparative method. It examines how platforms use ratings, personalised pay, gamification and automated deactivation to manage workers in ride-hailing, delivery and freelance work, and places this alongside the legislative and ballot-driven contest over worker status in California under Assembly Bill 5 and Proposition 22, and the European Union's 2024 Platform Work Directive, which introduces a rebuttable presumption of employment triggered by defined indicators of algorithmic direction and control.

The paper argues that algorithmic management does not eliminate employer control but relocates and conceals it, replacing a visible supervisor with an opaque system that performs the same disciplinary function while evading the doctrinal tests designed to detect it. The paper concludes that the law and workers alike need better tools to hold the invisible employer to account, starting with meaningful access for workers to the criteria used to evaluate and pay them and human review of automated decisions with serious consequences, such as deactivation.

Keywords: Algorithmic Management; Gig Economy; Control Test; Employment Classification; Platform Work

1 INTRODUCTION

Most workers who have a boss can picture that boss: a supervisor who assigns shifts, checks in during the day, and decides whether a raise or a warning is deserved. Millions of gig workers today have no such person to picture. The driver for a ride-hailing app, the courier for a food-delivery platform, and the freelancer on an online marketplace are all managed, disciplined, and paid by software rather than by a human being who watches them work. This paper calls that software the *invisible employer*.

Scholars use the term *algorithmic management* to describe this system: the use of software algorithms to assign tasks, monitor performance, evaluate quality, set pay, and discipline or terminate workers, largely without direct human supervision.[1] Companies market this arrangement as "flexible" and "entrepreneurial," since workers can log on and off whenever they like. Yet researchers who have studied these

[1] Alex Rosenblat & Luke Stark, *Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers*, 10 INT'L J. COMM. 3758, 3758–60 (2016).

platforms closely have found that algorithmic systems often exercise a form of control that is just as powerful as, and in some ways more invasive than, traditional supervision — it is simply harder to see.[2]

The stakes of understanding this shift are not merely academic. Tens of millions of people worldwide now earn some or all of their income through app-based platforms, and the number continues to grow as companies in retail, warehousing, and even office work experiment with similar tools. If algorithmic management really does function as a form of control comparable to traditional supervision, then treating platform workers as fully independent business owners — free of any employer at all — may rest on a legal fiction. Understanding exactly how the software controls workers is therefore a necessary first step before deciding how the law should respond.

This paper explains what algorithmic management is, traces how it grew out of older methods of workplace control, describes the specific tools platforms use to manage workers, illustrates those tools with examples from ride-hailing, delivery, and freelance platforms, and considers what this means for workers and for the law. Throughout, the goal is to make a complex and fast-moving academic literature accessible without oversimplifying the underlying research. The paper does not argue for a single legal fix; instead, it tries to show, clearly and with evidence, why the question of who — or what — is managing gig workers has become so difficult to answer.

2 WHAT IS ALGORITHMIC MANAGEMENT?

Algorithmic management refers to a bundle of practices in which data collection and computer algorithms replace many of the functions that a human manager used to perform.[3] Instead of a supervisor assigning the next job, a dispatching algorithm decides which worker receives which task, based on location, rating, acceptance history, and other data points the worker usually cannot see. Instead of a manager observing performance, the platform tracks metrics such as speed, cancellation rates, and customer ratings, and feeds them into a score that determines future opportunities.

Four features distinguish algorithmic management from ordinary workplace software. First, it is *continuous*: data is collected on nearly every action a worker takes, not just at scheduled review periods. Second, it is *opaque*: the exact rules that translate data into rewards or penalties are rarely disclosed to workers, and platforms frequently change them without notice.[4] Third, it is *adaptive*: the system can be tuned in real time for individual workers, meaning two people doing the same job may be paid or assigned differently based on hidden calculations. Fourth, it operates at a *scale* that no team of human managers could match, allowing a single platform to coordinate the work of millions of people simultaneously. Shoshana Zuboff has argued that this capacity to extract behavioral data and convert it into a tool of prediction and control is part of a broader economic logic she calls "surveillance capitalism," in which human experience itself becomes raw material for profit.[5]

It is worth pausing on why the *combination* of these four features matters more than any single one of them. Employers have monitored workers for as long as employment has existed, and older technologies

[2] Katherine C. Kellogg, Melissa A. Valentine & Angèle Christin, *Algorithms at Work: The New Contested Terrain of Control*, 14 ACAD. MGMT. ANNALS 366, 366–70 (2020).

[3] Kellogg, Valentine & Christin, *supra* note 2, at 371–75.

[4] Alex J. Wood, Mark Graham, Vili Lehdonvirta & Isis Hjorth, *Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy*, 33 WORK, EMP. & SOC'Y 56, 57 (2019).

[5] SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 201–05 (2019).

— time clocks, GPS trackers, call-center scripts — already automated pieces of supervision long before smartphone apps existed. What is genuinely new is the integration of monitoring, evaluation, task assignment, and pay-setting into a single, centrally designed system that touches almost every part of a worker's day and that can be updated instantly and silently by the company that owns it. It is this combination — not surveillance alone, and not automation alone — that researchers have in mind when they describe algorithmic management as a genuinely new "terrain of control."[6]

3 From the Foreman to the Algorithm: A Short History of Workplace Control

Employers have always looked for ways to manage labor efficiently. In the early twentieth century, scientific management ("Taylorism") broke jobs into small, measurable tasks that a foreman could time and inspect. Later in the century, bureaucratic organizations relied on written rules, job descriptions, and layers of middle management to standardize behavior across large workforces. What changed with the rise of digital platforms was not the desire to control labor, but the *tool* used to do it. Ride-hailing and delivery platforms replaced the foreman's stopwatch and the manager's rulebook with a smartphone app that watches, scores, and instructs the worker directly.[7]

Researchers Katherine Kellogg, Melissa Valentine, and Angèle Christin describe this as a shift to "algorithms at work" that perform six management functions once handled by people: restricting, recommending, recording, rating, replacing, and rewarding workers.[8] What makes this shift significant is not merely technical. Because the algorithm is embedded in the app that workers must use to find and complete jobs, it can enforce compliance instantly and automatically, without the delays, negotiations, or discretion that came with a human supervisor.

4 The Toolkit of Invisible Control

Although platforms differ, most rely on a similar set of tools to manage their workforce. Understanding these tools helps explain why algorithmic management feels different from ordinary employment, even though the underlying goal — getting work done efficiently and cheaply — is not new.

4.1 Ratings and Reputation Scores

Customer ratings, often on a five-point scale, are aggregated into a worker's overall score. Workers who fall below a threshold can be suspended or permanently removed from the platform ("deactivated"), frequently without a clear explanation or a meaningful appeal process. Hatim Rahman's research on a freelance labour platform found that because the criteria behind these scores were kept opaque and changed without warning, workers experienced what he calls an "invisible cage": they felt controlled by the evaluation system yet could not figure out how to align their behaviour with its shifting standards.[9] Some workers responded by obsessively experimenting with their profiles and work habits to try to improve their scores; others, especially those less dependent on the platform for income, simply gave up

[6] Kellogg, Valentine & Christin, *supra* note 2, at 366–68.
[7] Rosenblat & Stark, *supra* note 1, at 3761–63.
[8] Kellogg, Valentine & Christin, *supra* note 2, at 376–80.
[9] Hatim A. Rahman, *The Invisible Cage: Workers' Reactivity to Opaque Algorithmic Evaluations*, 66 ADMIN. SCI. Q. 945, 946–48 (2021).

trying to influence a system they could not understand.[10] In a later book-length study, Rahman shows that this dynamic persists even among highly skilled, in-demand freelancers, suggesting that opacity — not just a lack of bargaining power — is doing much of the controlling work.[11]

4.2 Personalized and Dynamic Pay

Rather than paying every worker the same rate for the same task, many platforms calculate pay individually, using data about each worker's history, location, and even how likely that worker is to accept or reject a given offer. Legal scholar Veena Dubal calls this practice "algorithmic wage discrimination": workers performing identical labour may be paid different amounts based on opaque, constantly shifting formulas designed to keep them working just long enough to hit a target.[12] Drivers Dubal interviewed compared the experience to gambling, describing a system of intermittent bonuses and unpredictable fares that resembles a slot machine more than a paycheck.[13] Because the formulas are proprietary, workers cannot verify whether the pay they receive is fair, and they cannot easily compare their pay to that of coworkers doing the same job.[14]

4.3 Gamification and Behavioral Nudges

Platforms frequently use game-like features — badges, streaks, progress bars, and surge-pricing alerts — to encourage workers to keep working during periods of high demand or to accept less desirable jobs. Mareike Möhlmann and Lior Zalmanson's study of Uber drivers found that such nudges created real tension: drivers felt pressure to follow the platform's suggestions even when they were not required to, because ignoring them seemed to lower future earning opportunities.[15] In response, drivers developed their own countertactics, such as coordinating mass log-offs to trigger surge pricing, illustrating that algorithmic control, however powerful, is not absolute; workers routinely find ways to "game" the very systems built to manage them.[16]

4.4 Automated Discipline and the Absent Manager

Perhaps the starkest feature of algorithmic management is automated deactivation: a worker can lose access to a platform, and therefore to their income, through an automated decision with little or no human review. Because there is no manager to appeal to in person, workers often describe the experience of being disciplined by an algorithm as uniquely disorienting — there is no one to argue with, and no clear standard to point to in one's defence.[17]

---

[10] *Id.* at 949–52.

[11] HATIM A. RAHMAN, INSIDE THE INVISIBLE CAGE: HOW ALGORITHMS CONTROL WORKERS 12–18 (2024).

[12] Veena Dubal, *On Algorithmic Wage Discrimination*, 123 COLUM. L. REV. 1929, 1934–38 (2023).

[13] *Id.* at 1961–64.

[14] *Id.* at 1965–68.

[15] Mareike Möhlmann & Lior Zalmanson, *Hands on the Wheel: Navigating Algorithmic Management and Uber Drivers' Autonomy*, *in* PROCEEDINGS OF THE 2017 INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS (ICIS) 1, 5–7 (2017).

[16] Wood, Graham, Lehdonvirta & Hjorth, *supra* note 4, at 62–65.

[17] Kellogg, Valentine & Christin, *supra* note 2, at 381–85.

5 Case Illustrations Across the Gig Economy

Ride-hailing is the most studied example. Rosenblat and Stark's nine-month study of Uber drivers found that although the company promoted an image of driver independence, its app exercised substantial indirect control — through suggested routes, performance messages, and rating thresholds — that shaped driver behaviour nearly as much as direct instruction would.[18] Drivers described receiving automated messages that read like gentle suggestions ("consider driving toward downtown"), but that functioned, in practice, as instructions, because ignoring them too often could mean fewer profitable ride offers in the future.[19]

Food-delivery platforms display a closely related pattern, with an added layer of physical risk. Couriers are typically paid per completed delivery rather than by the hour, and the algorithm that assigns orders can reward speed by offering more or better-paying jobs to couriers who accept quickly and complete deliveries fast. Because the same system that assigns work also tracks lateness and cancellations, couriers report feeling pressure to keep moving even in poor weather or unsafe traffic conditions, since slowing down risks a lower rating and, in turn, fewer future orders — the same rating-driven mechanism described among ride-hailing drivers, applied to workers who are also navigating city streets on bicycles or scooters.[20]

Remote "click-work" and freelance platforms show that algorithmic management is not limited to app-based transportation or delivery. A large cross-country study of remote gig workers by Alex Wood, Mark Graham, Vili Lehdonvirta, and Isis Hjorth found that algorithmic control was central to how online labour platforms operated worldwide, even though workers reported high levels of flexibility and autonomy in choosing tasks. The same features that produced flexibility — constant availability of new work, and rating-driven access to future jobs — also drove overwork, irregular hours, and social isolation.[21] The same team's related work on "embeddedness" found that remote gig workers often lacked the social and institutional supports — unions, local labour law protections, informal peer networks — that traditionally cushioned workers against employer power, leaving the platform's algorithm as one of the only consistent forces shaping their work lives.[22]

High-skilled freelance platforms complicate the assumption that algorithmic management only affects low-wage workers. Rahman's research on a professional freelance marketplace found that even highly educated, in-demand workers experienced the same disorientation from opaque rating criteria as workers in lower-skill gig jobs, suggesting that the structure of the platform — not the skill level of the worker — is what produces the "invisible cage."[23]

---

[18] Rosenblat & Stark, *supra* note 1, at 3770–72.
[19] *Id.* at 3763–65.
[20] Kellogg, Valentine & Christin, *supra* note 2, at 378–80.
[21] Wood, Graham, Lehdonvirta & Hjorth, *supra* note 4, at 58–61.
[22] Alex J. Wood, Mark Graham, Vili Lehdonvirta & Isis Hjorth, *Networked but Commodified: The (Dis)Embeddedness of Digital Labour in the Gig Economy*, 53 SOCIOLOGY 931, 935–40 (2019).
[23] Rahman, *supra* note 9, at 953–58.

6 CONSEQUENCES FOR WORKERS

The consequences of algorithmic management fall into three broad categories: an autonomy paradox, economic insecurity, and weakened collective power. Each is worth examining in turn, because each points toward a different kind of remedy.

First, there is what researchers call the *autonomy paradox*: platforms advertise freedom and flexibility, yet the same systems that grant workers the freedom to log on whenever they like also quietly steer, rank, and discipline them. Workers often experience this as freedom *with strings attached* — they can say no to a task, but saying no too often can quietly lower their future earnings or standing.[24]

Second, opaque evaluation and pay systems create economic insecurity. Because workers cannot verify how their pay or their rating is calculated, they cannot reliably predict their income or plan their finances, a problem Dubal documents in detail through interviews describing the psychological toll of what drivers call a "casino" style of pay.[25] This unpredictability is compounded by the fact that legal protections designed for traditional employees — minimum wage guarantees, unemployment insurance, protection from arbitrary firing — generally do not apply to workers classified as independent contractors.[26]

Third, algorithmic opacity undermines workers' ability to organize collectively or seek redress. When workers cannot identify why they were paid a certain amount or deactivated from a platform, they struggle to build the kind of shared, verifiable grievances that have historically fueled labor organizing.[27] If two drivers cannot easily confirm that they were paid differently for comparable trips, it becomes far harder to prove a pattern of unfair treatment, let alone to bargain collectively over it. Individualized, opaque pay formulas thus do more than confuse individual workers — they can quietly weaken the shared information that collective action has traditionally depended on.[28]

There is also a less-studied but increasingly documented psychological dimension to these consequences. Workers describe a persistent, low-grade anxiety about being watched and scored at all times, along with genuine confusion about whether a bad day, a difficult customer, or simple bad luck will translate into a lasting mark against them. Rahman's interviews capture this directly: workers who had experienced a sudden, unexplained drop in their rating described ongoing worry about when — or whether — it might happen again, even after their scores recovered.[29] This kind of chronic uncertainty, produced not by the difficulty of the work itself but by not knowing the rules that govern it, is one of the clearest ways in which algorithmic management differs in character from traditional supervision, however imperfect that older system may also have been.

7 LEGAL AND POLICY RESPONSES

Legal systems have begun to grapple with algorithmic management, primarily through the lens of worker classification and, more recently, through direct regulation of algorithmic pay and evaluation practices. Two distinct strategies have emerged: reclassifying workers so that existing employment law applies to

[24] Dubal, *supra* note 12, at 1944–52.
[25] *Id.* at 1961–64.
[26] Veena B. Dubal, *Wage Slave or Entrepreneur?: Contesting the Dualism of Legal Worker Identities*, 105 CALIF. L. REV. 65, 68–72 (2017).
[27] *Id.* at 101–06.
[28] Dubal, *supra* note 12, at 1970–75.
[29] Rahman, *supra* note 9, at 949–52.

them, and writing new rules that target algorithmic management directly, regardless of how a worker is classified.

In the United States, the central legal battle has been over whether gig workers are "employees," entitled to protections such as minimum wage and overtime, or "independent contractors," who are not. Veena Dubal's scholarship on this dualism was cited by the California Supreme Court in *Dynamex Operations West, Inc. v. Superior Court*, which adopted a stricter "ABC test" presuming that most workers are employees unless the hiring company proves otherwise.[30] The California legislature soon codified this standard in Assembly Bill 5, but voters later approved Proposition 22, which carved out an exemption for app-based drivers, allowing companies to continue classifying them as independent contractors while offering certain limited benefits.[31] Dubal has argued that this compromise effectively legalized algorithmic wage discrimination by allowing companies to set individualized pay without the transparency obligations that would apply to a traditional wage system.[32]

The European Union has taken a more direct regulatory approach. The Platform Work Directive, adopted in 2024, requires platforms to disclose the parameters used in automated monitoring and decision-making systems, prohibits certain forms of automated dismissal without human review, and creates a rebuttable presumption of employment status for platform workers who meet specified criteria.[33] Scholars view this as one of the first serious legislative attempts to regulate algorithmic management directly, rather than relying solely on older employment-classification doctrines.[34]

Beyond formal law, some scholars argue that transparency requirements alone will not solve the deeper problem, because even a fully disclosed algorithm can still concentrate enormous power in the hands of a platform that controls the data.[35] Others emphasize that workplace algorithms are not neutral technical artifacts but embody specific managerial choices, meaning that meaningful reform will likely require worker input into how these systems are designed, not merely after-the-fact disclosure of how they operate.[36]

These two strategies — classification and direct regulation — are not mutually exclusive, and several jurisdictions are now experimenting with both at once. Reclassification extends an existing, well-understood body of protections to platform workers, but it does so at the cost of long, uncertain litigation and the risk that companies will simply restructure their businesses to avoid the new rules, as some argue happened with Proposition 22. Direct regulation of algorithmic management, by contrast, can reach workers regardless of how they are classified, but it requires regulators to understand technical systems that companies have strong incentives to keep opaque. The Platform Work Directive's requirement that companies disclose the logic behind automated decisions is, in this sense, a test of whether legal transparency mandates can meaningfully open up what has so far been a genuinely invisible process.[37]

---

[30] Dynamex Operations W., Inc. v. Superior Court, 416 P.3d 1 (Cal. 2018); *see also* Dubal, *supra* note 26, at 72 (identifying the "ABC test" later adopted by the California Supreme Court).

[31] Cal. Assemb. B. 5, 2019–2020 Leg., Reg. Sess. (Cal. 2019) (codifying the ABC test); Cal. Prop. 22 (2020) (exempting app-based drivers from AB 5's employee classification).

[32] Dubal, *supra* note 12, at 1985–90.

[33] Directive 2024/2831 of the European Parliament and of the Council of 23 October 2024 on Improving Working Conditions in Platform Work, 2024 O.J. (L) 2831.

[34] Rahman, *supra* note 11, at 190–97.

[35] Zuboff, *supra* note 5, at 8–12, 351–56.

[36] Kellogg, Valentine & Christin, *supra* note 2, at 390–95.

[37] Directive 2024/2831 of the European Parliament and of the Council of 23 October 2024 on Improving Working Conditions in Platform Work, 2024 O.J. (L) 2831.

8 CONCLUSION

Algorithmic management has not eliminated the employer — it has made the employer harder to see. The dispatching algorithm, the rating system, and the dynamic pay formula now perform many of the functions once carried out by a human supervisor, but they do so with less transparency, less accountability, and, in many legal systems, less regulation. The research reviewed in this paper shows a consistent pattern: platforms offer workers real flexibility and autonomy, but they pair that flexibility with opaque, constantly shifting systems of evaluation and pay that function as a genuine, if invisible, form of workplace control.

As algorithmic management spreads beyond ride-hailing and delivery into warehousing, retail, and even white-collar work, the questions raised by this research become more urgent. Whether the answer lies in reclassifying workers as employees, mandating algorithmic transparency, giving workers a genuine voice in how these systems are designed, or rethinking labour law altogether, one conclusion is hard to avoid: an employer that cannot be seen is still an employer, and the law and workers alike will need better tools to hold it to account.

The literature reviewed here suggests three practical starting points for that project. First, workers need meaningful access to the criteria used to evaluate and pay them, not merely a general assurance that an algorithm is being used fairly. Second, automated decisions with serious consequences — deactivation chief among them — should be subject to some form of human review before they take effect, not only after a worker has already lost income or access to a platform. Third, because individual workers rarely have the resources to challenge an opaque system on their own, researchers, regulators, and worker organizations will likely need to keep collecting the kind of empirical evidence this paper has drawn on, so that the invisible employer can be studied, understood, and, where necessary, held accountable, one dataset at a time.

* * *

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.16

CHAPTER 16

From Corporate Control to Algorithmic Control: Rethinking Directors' Duties and Liability in AI-Driven Corporations

K G SAI SAMYUKHTHA*

Corporate governance has traditionally assumed that important company decisions are made by human directors, managers, and officers who can be questioned about the reasons for their choices. Artificial intelligence is changing that assumption. AI systems are now used to analyse markets, assess risk, screen employees, detect fraud, forecast demand, recommend investments, and in some settings take actions with little or no human intervention. The legal problem is not simply whether a company may use AI. It is whether existing rules on directors' duties can still allocate responsibility when an important corporate decision is partly or largely produced by an algorithm.

This paper examines that problem mainly through Indian company law, with comparative reference to Delaware, the United Kingdom, the European Union, the OECD and the United States National Institute of Standards and Technology (NIST). Section 166 of the Companies Act, 2013 requires directors to act in good faith, exercise due and reasonable care, skill and diligence, and use independent judgment. These duties remain relevant when AI is used. The difficulty is that the traditional idea of a director as the person who gathers information and then exercises judgment does not fit neatly where an AI system filters information, recommends an outcome, or executes the decision itself.

The paper argues that AI should not become a legal excuse for directors, but directors should also not become automatic guarantors of every error made by a system they did not design or operate. Liability should instead be linked to the degree of AI autonomy, the importance of the decision, the quality of human oversight, the foreseeability of the risk, and the ability of the company and its directors to intervene. The paper proposes an AI Corporate Responsibility Framework with four levels of AI use: assistive, delegated, supervised autonomous, and high-autonomy decision systems. For each level, it proposes a different combination of board oversight, documentation, testing, human review, vendor responsibility, and director liability.

The central conclusion is that the duty of directors should evolve into a clear duty of AI oversight. This duty would require directors to understand the material AI systems used by the company, identify significant risks, establish suitable controls, monitor performance, keep records of important AI-supported decisions, and intervene when a system presents a serious or foreseeable risk. Such a framework protects the basic principle of director accountability while recognising that modern corporate decisions may be produced through a network of directors, employees, software providers, data sources, and autonomous systems.

Keywords: Artificial Intelligence; Algorithmic Decision-Making; Corporate Governance; Directors' Duties; Director Liability; AI Oversight; Corporate Responsibility

1 RESEARCH PROBLEM

The research problem is the mismatch between a legal system built around human corporate judgment and corporate systems that increasingly rely on algorithmic recommendations and autonomous action.

* VIT School of Law, Vellore Institute of Technology, Chennai, Tamil Nadu, India.

Existing company law can usually identify the company, its directors, officers, employees, and service providers. It is less clear how responsibility should be divided when an AI system materially changes the information available to a board or acts without a director making the immediate decision. The problem becomes sharper when an AI system makes a decision that is commercially reasonable in ordinary circumstances but harmful in a particular case, or when the board knew that the system was unreliable but continued to use it.

1.1 Problem Statement

The present legal framework does not clearly state how director duties should operate across different degrees of AI autonomy. If directors are held responsible for every AI error, the law may discourage useful technology and impose an unreasonable burden on individuals. If directors can avoid liability by saying that an algorithm made the decision, the basic purpose of corporate oversight is weakened. The legal gap is therefore not a simple absence of AI legislation. It is an allocation problem: which actor should bear responsibility for a harmful AI-supported corporate decision, and what level of oversight should the law reasonably expect from directors?

1.2 Research Objectives

- To examine how directors' duties under Indian company law apply when AI systems are used in corporate decision-making.
- To identify the circumstances in which directors may be liable for AI-related corporate harm and the circumstances in which liability should fall primarily on other actors.
- To compare the Indian position with selected approaches in Delaware, the United Kingdom, the European Union, and international AI governance frameworks.
- To assess whether existing duties of care, good faith, independent judgment, and oversight are sufficient for AI-driven corporate governance.
- To propose a practical AI Corporate Responsibility Framework that connects the level of AI autonomy with the level of human oversight and legal responsibility.

1.3 Research Questions

- How do existing directors' duties apply when an AI system materially influences or makes a corporate decision?
- When should a director be held liable for harm caused by an AI system used by the company?
- Can a director rely on an AI system in the same way that a director may rely on human officers, experts, or professional advisers?
- What can Indian corporate law learn from Delaware oversight doctrine, UK corporate governance, the EU AI Act, and international AI risk frameworks?
- What legal and governance structure can distribute responsibility between directors, companies, employees, and AI providers without weakening director accountability?

1.4 Methodology

This paper uses doctrinal and comparative legal research. The doctrinal part examines statutory provisions, judicial decisions, and established principles concerning directors' duties, oversight, delegation, and corporate responsibility. The principal Indian source is the Companies Act, 2013, especially section 166. The comparative analysis considers Delaware fiduciary and oversight law, the UK Corporate Governance Code, the EU Artificial Intelligence Act, the OECD AI Principles, and the NIST AI Risk Management Framework. Recent academic writing on AI and corporate governance is used to identify emerging arguments and areas of disagreement. The paper does not claim that any of these frameworks creates a complete legal answer to AI-driven corporate liability. Instead, it uses them to build a workable model for Indian corporate governance.

The method is also normative. After identifying what existing law can already do, the paper asks what additional governance requirements would make the allocation of responsibility fair and workable. The proposed framework is therefore not presented as current Indian law. It is a reform proposal designed to fit within the existing logic of directors' duties.

2 THE BOARDROOM MEETS THE ALGORITHM: INTRODUCTION

Company law has always depended on a practical fiction: a company is a separate legal person, but people must act for it. The board of directors is one of the main human institutions through which that legal person makes choices. Directors decide strategy, approve major transactions, supervise management, monitor risk, and protect the company's interests. Indian company law reflects this model by imposing duties on directors as individuals. Section 166 of the Companies Act, 2013 requires a director to act in good faith, exercise due and reasonable care, skill and diligence, and exercise independent judgment.[1] The assumptions behind these duties are now under pressure. A company may use AI to decide which customers receive credit, which transactions should be blocked, how much inventory should be purchased, which employees should be shortlisted, or which risks deserve management attention. In more advanced systems, software may not merely recommend an outcome. It may take an action automatically after receiving data from other systems. The human director may therefore supervise a process rather than make each decision directly.[2]

The change is important because corporate law generally allocates responsibility through human roles. A director can be asked what information was considered, why a particular course was chosen, whether alternatives were examined, and whether reasonable precautions were taken. An AI system does not fit comfortably into that model. It has no legal personality as a director, no fiduciary duty to the company, and no independent legal judgment in the corporate-law sense. It is a tool or system deployed by human and corporate actors. That does not mean the tool is legally irrelevant. Its design, data, level of autonomy, contractual terms, monitoring arrangements, and known limitations may all affect whether the human actors acted reasonably.

This is why the question should not be framed simply as "Who is liable when AI makes a mistake?" The better question is "Who made the conditions under which the AI was allowed to act, and what safeguards were reasonably required?" This moves the analysis from the final algorithmic output to the

[1] Companies Act, 2013, § 166(2)–(3).

[2] *See* Deirdre Ahern, *Directors' Duties in the Age of Agentic Artificial Intelligence*, CAMBRIDGE FORUM ON AI: L. & GOVERNANCE (2026).

governance process that produced it. Recent scholarship increasingly takes this approach, treating AI oversight as an extension of existing fiduciary and care duties rather than as a wholly new area of law.[3] The Indian position is especially interesting because section 166 already contains language that can accommodate technological change. The requirements of care, skill, diligence, and independent judgment are open-textured. They can respond to new business practices without requiring a separate statutory duty for every technology. The harder question is how far those words should extend when a director reasonably relies on a complex system whose internal operation the director cannot fully understand.[4]

This paper therefore starts from a modest proposition: AI does not remove the director from the legal picture. Instead, it changes the director's task. A director who once had to review information personally may now have to govern the information system that produces it. A director who once approved a decision may now have to approve the conditions under which an automated system is permitted to make decisions. In that sense, the central shift is from direct control to supervised control.

3 FROM HUMAN JUDGMENT TO ALGORITHMIC DECISION-MAKING

3.1 The Changing Corporate Decision Process

AI in a corporation exists on a spectrum. At one end, a system may simply organise information or produce a report. The director remains the clear decision-maker. At the other end, an AI agent may receive a goal, gather information, select between options, communicate with other systems, and execute actions. Between these points are systems that recommend decisions, systems that make routine decisions subject to human review, and systems that can act unless a human intervenes.

This spectrum matters legally. The more autonomous the system, the weaker the argument that the director personally made the immediate decision. At the same time, greater autonomy creates a stronger need for governance before the system is deployed. A board that deliberately gives an AI system authority to make high-impact decisions cannot reasonably treat the system as an ordinary piece of office software.

Corporate governance literature already treats boards as bodies responsible for strategy, risk, compliance, and oversight. The same reasoning applies when AI becomes part of those functions. The UK Corporate Governance Code 2024, for example, places strong emphasis on board leadership, division of responsibilities, audit, risk, and internal control. It does not create a special "AI director" role, but its structure is capable of covering AI risk as part of corporate risk and control.[5] The NIST AI Risk Management Framework similarly treats AI risk management as a continuous process organised around the functions of govern, map, measure, and manage. Although NIST is not a corporate law statute, its structure is useful because it shows how a company can convert general responsibility into repeatable governance practices.[6]

3.2 Why Algorithmic Error Is Different

An ordinary human error and an algorithmic error can look similar at the end of the process, but their causes may be very different. A human manager may misunderstand a document. An AI system may

[3] *See* Martin Petrin, *AI Accountability from the Outside In: Corporate Liability, AI Governance, and Managerial Duties* (ECGI Working Paper No. 935, 2026).
[4] Companies Act, 2013, § 166(3).
[5] FINANCIAL REPORTING COUNCIL, UK CORPORATE GOVERNANCE CODE 2024 (Jan. 22, 2024).
[6] ELHAM TABASSI, ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), NIST AI 100-1 (2023).

produce an inaccurate result because its training data was incomplete, because a model changed after deployment, because a data pipeline was corrupted, because a proxy variable created discriminatory outcomes, or because the system was used outside the setting for which it was tested. A director may not have caused the technical failure, but the director may still have been responsible for choosing the system, approving its use, setting its limits, or failing to investigate warning signs.

That distinction suggests a process-based approach to liability. The court should ask what the board knew, what the board could reasonably have known, what the company was using the system for, what controls existed, whether the system had been tested, whether there was a meaningful human override, and whether the risk was material to the company's business. This approach is closer to the traditional duty of care than a rule that simply attaches liability to the person whose name appears on the board.

4 What Existing Law Already Says About Directors' Duties

4.1 India: Section 166 and the Human Director

Section 166 is the starting point for an Indian analysis. A director must act in good faith to promote the objects of the company and in the best interests of the company, its members, employees, shareholders, the community, and the protection of the environment. The director must also exercise due and reasonable care, skill and diligence and independent judgment.[7] The language is broad enough to cover AI-related conduct without treating AI as a legal person. If a board approves an AI system for a material function, the directors can be expected to consider whether the system is suitable for that purpose, whether the company can monitor it, and whether important decisions can be reviewed or reversed.[8]

Independent judgment is particularly important. Reliance on technology does not automatically mean that a director has given up independent judgment. Directors routinely rely on financial statements, lawyers, auditors, engineers, consultants, and management reports. The legal issue is whether the reliance was reasonable. If the director treats an AI output as automatically correct despite obvious limitations, the problem is not the use of AI itself. The problem is the absence of independent judgment.

The same reasoning applies to diligence. Due diligence in an AI-dependent company may include asking what data the system uses, what its known failure modes are, how often it is tested, who can change its parameters, what happens when it produces an abnormal result, and whether the company has retained enough human expertise to challenge it. A director need not become a software engineer. But a director cannot reasonably remain unaware of a technology that performs a material corporate function.

4.2 Delegation and Reliance

Corporate law accepts that directors cannot personally perform every task. They govern through executives, committees, employees, and outside advisers. This creates a useful comparison with AI. If a director may rely on a chief financial officer or an external auditor in appropriate circumstances, it would be strange to say that reliance on a technological system is always forbidden. But AI reliance differs in one important respect: the system may not be a legally accountable professional with a duty to the company. A director therefore has to examine the chain behind the output.

---

[7] Companies Act, 2013, § 166(2)–(3).
[8] *Id.* § 166(3).

A sensible legal rule is that reliance should be stronger where the AI system is narrow, tested, explainable enough for its purpose, and subject to effective human review. Reliance should be weaker where the system is opaque, self-modifying, connected to external systems, capable of taking irreversible action, or used for decisions involving substantial legal, financial, or human consequences.

4.3 The Comparative Lesson from Delaware

Delaware provides a useful comparison because its corporate law places substantial weight on board process and oversight. The Caremark line of cases developed the idea that directors may face liability for a sustained failure to exercise oversight, particularly where they fail to establish or monitor systems designed to bring important risks to the board's attention.[9] In Marchand v. Barnhill, the Delaware Supreme Court stressed that a board-level monitoring system can be especially important when a risk is central to the company's business. The lesson for AI is not that every algorithm requires a Caremark claim. It is that boards cannot treat a central risk as somebody else's problem when they have the power and responsibility to oversee it.[10]

The comparison is useful because AI can create both ordinary business risks and compliance risks. A pricing model that loses money may primarily raise a business judgment issue. An automated system that repeatedly violates privacy law, discriminates against protected groups, or creates serious safety risks may instead raise an oversight issue. The legal analysis should therefore distinguish between a bad result and a governance failure.

5 WHO SHOULD PAY WHEN THE ALGORITHM GOES WRONG?

5.1 The False Choice between Director Liability and No Liability

The strongest temptation in AI liability debates is to choose one responsible party. If the algorithm caused the harm, some may say the software developer should be liable. Others may say that the company owns the system and must therefore bear the loss. Still others may argue that the directors approved the system and should be personally liable. Each position is too simple when used alone.

A corporation can be responsible even when no individual director is personally at fault. A vendor can be responsible for defective software or false representations. Employees can be responsible where they knowingly misuse a system. Directors can be responsible where they fail to exercise the care or oversight required by their role. These forms of responsibility can coexist.

5.2 The Allocation Factors

- Control: Who selected the system, set its purpose, and decided how much authority it had?
- Knowledge: What did the relevant actor know about the system's limitations, failures, and risks?
- Foreseeability: Was the harm a reasonably foreseeable consequence of the system's use?
- Materiality: Was the AI system involved in a decision that was important to the company, its finances, compliance, or stakeholders?

[9] Stone v. Ritter, 911 A.2d 362, 370 (Del. 2006).
[10] Marchand v. Barnhill, 212 A.3d 805, 821 (Del. 2019).

- Autonomy: Did the system merely advise a human, or did it act without meaningful human approval?
- Intervention: Could a human reasonably detect, stop, or reverse the system's action?
- Documentation: Did the company keep records showing how the system was tested, monitored, and used?
- Vendor responsibility: Did the supplier make material representations about the system's accuracy, safety, or compliance, and were those representations reasonable to rely upon?

These factors help separate a director's personal liability from corporate liability. Suppose a board approves a widely used commercial fraud-detection system after obtaining expert assurance, conducting testing, setting conservative thresholds, and requiring human review of high-risk cases. If a rare and unforeseeable model failure later causes loss, personal director liability should be difficult to establish. Now change the facts: the board knows the system has repeatedly generated false results, removes the human review to save costs, and gives the system authority to block customers automatically. A later harm would be much more difficult to defend as a mere technological accident.

5.3 The EU Approach and Its Relevance to Corporate Governance

The European Union's AI Act takes a risk-based approach and places specific obligations on providers and deployers of certain AI systems. The Act does not replace company law or create a general rule that directors are personally liable for AI outcomes. Its importance for corporate governance lies in its emphasis on risk classification, human oversight, technical documentation, monitoring, and accountability.[11] The EU approach therefore supports a broader principle: legal responsibility should increase with the risk and autonomy of the system. This is compatible with, rather than contrary to, the Indian duty of care.[12]

6 Comparative Models for AI Oversight

6.1 United Kingdom: Board Responsibility through Risk and Control

The UK Corporate Governance Code 2024 does not create an AI-specific director duty. Its significance is structural. The Code treats board leadership, division of responsibility, audit, risk, internal control, and reporting as core governance areas. From 2026, Provision 29 requires boards within its scope to make a declaration concerning the effectiveness of material internal controls.[13]

For AI-driven corporations, this suggests that AI should be treated as part of the company's internal control environment when it performs a material function. A board should know where important AI systems are used, who owns each system internally, what risks have been identified, and what controls exist. The UK model is valuable because it does not depend on creating a new category of corporate officer for every new technology.

---

[11] Regulation (EU) 2024/1689 (Artificial Intelligence Act), arts. 1, 9–15, 2024 O.J. (L 1689) 1.

[12] *Id.*

[13] Financial Reporting Council, UK Corporate Governance Code 2024, Provision 29 and related guidance (2024).

6.2 OECD: Accountability and Human Oversight

The OECD AI Principles, updated in 2024, emphasise transparency and explainability, robustness and security, and accountability. They also call for human agency and oversight appropriate to the context and for traceability across the AI system lifecycle.[14]

These principles translate well into corporate governance. A director cannot oversee an AI system if the company does not know what data enters it, what model is being used, who can change it, what outputs it produces, and what happens after an output is generated. Traceability is therefore not only a technical concept. It is also evidence for corporate accountability.

6.3 NIST: A Practical Risk Management Model

The NIST AI RMF 1.0 uses four broad functions: Govern, Map, Measure, and Manage. It is voluntary and not a legal standard, but it provides a practical structure for converting broad responsibility into procedures.[15]

In a corporate setting, "Govern" can mean assigning board and management responsibility. "Map" can mean identifying AI systems, their uses, affected stakeholders, data sources, and foreseeable risks. "Measure" can include testing accuracy, bias, security, robustness, and performance. "Manage" can include mitigation, human review, incident response, suspension, and retirement. The value of such a model is that it creates a record of what the company actually did.

6.4 The Indian Gap

India has a strong general corporate law framework, but it does not yet have a single corporate-governance statute that sets out a detailed board protocol for AI systems. This does not mean there is no law. Section 166, securities obligations, data protection rules, consumer law, sectoral regulation, contractual law, and general principles of negligence and corporate responsibility can already apply depending on the facts. The gap is more specific: there is no settled method for translating the general duty of care into expectations for AI selection, testing, monitoring, and intervention.

7 A PROPOSED AI CORPORATE RESPONSIBILITY FRAMEWORK

The paper proposes a four-level AI Corporate Responsibility Framework. The framework is not intended to make AI itself a legal person. It is intended to classify the relationship between the company, its directors, and the AI system so that the level of oversight matches the level of risk.

7.1 Level I — Assistive AI: The Director Decides

At Level I, AI provides information, summaries, forecasts, or recommendations, but a human director or officer makes the actual decision. Examples include board briefing tools, financial analysis, market research, and document review. Director responsibility remains close to the ordinary duty of care. The director should assess whether the information is reliable enough for the decision and should not blindly accept an output merely because it is computer-generated.

---

[14] OECD, OECD AI PRINCIPLES (updated 2024).

[15] ELHAM TABASSI, ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), NIST AI 100-1 (2023).

Controls should include basic source checking, disclosure that AI was used where appropriate, and a rule that material decisions cannot be justified solely by saying "the AI recommended it."

7.2 Level II — Delegated AI: The System Decides Routine Matters

At Level II, AI is authorised to make defined decisions within limits. Examples may include routine fraud flags, inventory reordering, standard customer service actions, or low-value transaction approvals. The board does not make each decision, but it approves the system's authority.

Here the director's duty shifts toward system governance. The board should approve the purpose, limits, escalation rules, and monitoring arrangements. The company should maintain logs and periodically test whether the system remains within its approved purpose. If the board knows that the system regularly exceeds its limits and does nothing, director exposure should increase.

7.3 Level III — Supervised Autonomous AI: The System Acts, Humans Monitor

At Level III, an AI system can make material decisions and act without immediate approval, but a human remains responsible for monitoring and intervention. This level is likely to become more important as agentic systems become capable of carrying out multi-step tasks.

At this level, the company should have a named executive owner, board-level reporting for material systems, documented risk assessments, testing before deployment, ongoing monitoring, incident reporting, a meaningful override mechanism, and clear rules for suspension. Directors should receive enough information to understand the system's role and risk even if they do not understand its source code.

This approach is consistent with the broader direction of modern AI governance, which places weight on human oversight, traceability, risk management, and the ability to intervene when a system behaves in an unsafe or unexpected way.[16]

7.4 Level IV — High-Autonomy AI: The System Can Create Material Corporate Consequences

Level IV covers systems that can take significant financial, legal, operational, or stakeholder-facing actions with little human involvement. Examples could include an agent authorised to negotiate and enter contracts, allocate large amounts of capital, change material pricing, or make decisions with serious consequences for employees or customers.

These systems should face the strongest controls. The board should expressly approve their use. The company should conduct periodic independent audits, maintain detailed decision logs, test for failure and misuse, define prohibited actions, and ensure that the system can be stopped. Material changes to the system should trigger a fresh risk assessment rather than being treated as routine software maintenance.

7.5 The Proposed AI Oversight Duty

The four-level model supports a general AI oversight duty within the existing duty of care. The duty would contain five basic elements: identification, assessment, control, monitoring, and intervention.

- Identification — directors should know which AI systems perform material corporate functions.

[16] OECD, OECD AI PRINCIPLES (updated 2024); Regulation (EU) 2024/1689, arts. 9–15; NIST, AI RMF 1.0.

- Assessment — directors should consider the system's purpose, reliability, data, autonomy, legal risks, and foreseeable harms.
- Control — directors should ensure that authority is limited to what the company has approved and that responsibility is assigned.
- Monitoring — directors should receive meaningful reports about failures, unusual outputs, incidents, and material changes.
- Intervention — directors should ensure that the company can suspend, correct, or replace a system when serious risk appears.

7.6 When Should Directors Be Personally Liable?

Personal liability should not arise simply because an AI system produced a harmful result. The better test is whether the director breached the applicable duty in relation to the company's use of the system. Liability becomes stronger where the director approved a high-risk system without reasonable inquiry, ignored known failures, failed to establish basic controls, allowed an autonomous system to act beyond its approved purpose, or failed to respond to material warning signs.

Conversely, liability should be weaker where the company used a system for a legitimate purpose, performed reasonable testing, obtained competent advice, maintained suitable controls, monitored the system, and responded appropriately to incidents. Corporate law should leave room for genuine business risk. The purpose of a duty of care is not to make directors insurers against every loss.

7.7 Responsibility of AI Providers and Other Actors

The framework also recognises that directors do not control every part of an AI system. A software provider may control the model architecture; a data supplier may control training or input data; an employee may configure the system; and an outside consultant may make important representations about performance. Contracts should therefore identify responsibility for testing, security, updates, incident reporting, data quality, and material changes.

However, contractual allocation should not be used to erase the board's governance role. A director cannot avoid an internal corporate duty merely by signing a vendor agreement. The agreement can allocate operational responsibility, but the board still has to decide whether the company should rely on the system.

7.8 Documentation as a Legal Safeguard

Documentation should be treated as part of governance, not as paperwork added after an incident. For material AI systems, the company should retain a short record of the system's purpose, level of autonomy, risk assessment, testing, responsible persons, human-override arrangements, known limitations, major incidents, and significant changes.

This protects both the company and directors. If the system later causes harm, the record can show whether the board acted reasonably at the time. It also discourages hindsight-based judgments. A good governance record does not guarantee immunity, but it makes the decision-making process visible.

8 Conclusion: Keeping Human Accountability in an Automated Corporation

AI does not make directors' duties obsolete. It changes what responsible directorship looks like. The traditional director was expected to obtain information, assess it, ask questions, make a decision, and supervise implementation. In an AI-driven company, part of that work may be performed by a system. The director's responsibility therefore moves upward: from checking every individual output to governing the system that produces the outputs.

Indian company law already contains much of the legal foundation needed for this change. Section 166 requires good faith, care, skill, diligence, and independent judgment. Those duties are sufficiently flexible to apply to technology that did not exist when the Companies Act was enacted. What is missing is a clearer understanding of how those duties operate when AI is material, autonomous, and capable of producing decisions at scale.

Comparative developments support a process-based answer. Delaware's oversight cases show why boards must pay attention to important risks and maintain appropriate systems of oversight. The UK governance framework places risk and internal control within the board's responsibility. The EU AI Act uses a risk-based structure and requires governance measures for certain AI systems. The OECD Principles stress accountability, transparency, human oversight, robustness, and traceability. NIST offers a practical risk-management structure. None of these frameworks provides a complete Indian solution, but together they show that AI governance is becoming part of ordinary organisational responsibility.[17]

The proposed AI Corporate Responsibility Framework therefore rejects two extreme positions. The first is that directors should remain fully liable simply because they are directors. The second is that directors should escape responsibility because "the algorithm decided." Neither approach reflects how modern corporations actually operate. Responsibility should follow control, knowledge, foreseeability, materiality, autonomy, and the quality of oversight.

The framework's four levels—assistive, delegated, supervised autonomous, and high-autonomy—provide a simple way to connect the level of AI freedom with the level of corporate oversight. At low levels, ordinary professional judgment may be enough. As autonomy and risk increase, stronger testing, monitoring, documentation, board involvement, human override, and independent review should be required. The legal consequence is not automatic director liability. Instead, failure to meet the level of oversight reasonably expected for the system should become evidence in assessing whether a director breached the existing duty of care.

This approach also preserves the separate legal personality of the corporation. AI need not be made a director, officer, or legal person merely because it performs a task that once required a human. Making software a legal director would not solve the real problem because the system cannot bear fiduciary duties in the same way as a human office-holder. The more useful response is to keep human and corporate responsibility in place while recognising that decision-making is increasingly distributed across people, software, data, and vendors.

Ultimately, the central principle should be simple: a director should not be punished for using AI responsibly, but a director should not be permitted to use AI as a shield against responsibility. The law should ask whether the director understood the role of the system, assessed its risks, put suitable controls

[17] Stone v. Ritter, 911 A.2d 362, 370 (Del. 2006); Financial Reporting Council, UK Corporate Governance Code 2024; Regulation (EU) 2024/1689; OECD, OECD AI Principles (updated 2024); NIST, AI RMF 1.0.

in place, monitored material performance, and acted when warning signs appeared. If those steps were taken, an unexpected AI failure should not automatically become personal liability. If those steps were ignored, the fact that an algorithm produced the final output should not break the chain of accountability.

The future of corporate governance is therefore unlikely to be a choice between human control and machine control. It is more likely to be a system of human responsibility exercised through increasingly autonomous tools. The law should adapt to that reality without abandoning the reason directors have duties in the first place: someone must remain answerable for how the corporation chooses to act.

* * *

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.17

CHAPTER 17

Algorithmic Decision Making and Corporate Insolvency: The Future of Creditor Control in the Digital Decade

ASHUTOSH MANI PATHAK*

Artificial intelligence, machine learning and predictive analytics increasingly support credit monitoring, risk assessment and aspects of corporate insolvency administration. Existing research identifies emerging uses of artificial intelligence in tasks involving claims, investigations, asset management and financial analysis, while emphasising both potential efficiency gains and concerns regarding the accuracy of algorithmic outputs. These developments raise a distinct legal question within creditor-controlled insolvency systems: when creditor decisions are materially influenced by algorithmic recommendations, under what conditions should those decisions remain protected by the judicial restraint traditionally associated with commercial wisdom?[1]

This paper examines the relationship between algorithmic decision-support, creditor control and commercial judgment under Indian corporate insolvency law. It does not claim that Committees of Creditors presently delegate votes or resolution decisions to autonomous systems. Instead, it addresses the legally significant continuum between existing digital risk assessment and the increasing capacity of algorithmic tools to influence valuation, recovery analysis, viability assessment and comparison of resolution alternatives. The paper distinguishes routine automated execution from algorithmic decision support and autonomous decision making, arguing that legal scrutiny should correspond to the materiality of technological influence.

Using doctrinal analysis of the Insolvency and Bankruptcy Code, 2016 and commercial wisdom jurisprudence, together with a functional comparison of emerging European and Indian digital regulation principles, the paper evaluates the benefits and risks of algorithmically informed creditor control. The analysis identifies earlier distress detection, improved information processing and administrative efficiency as potential benefits, but highlights opacity, automation bias, deficient data, unequal technological capacity and fragmented responsibility as threats to procedural legitimacy. The European Union Artificial Intelligence Act and the Reserve Bank of India's Digital Lending Directions, 2025 are used as regulatory analogies rather than as rules directly governing insolvency voting.[2]

The paper argues that reliance upon algorithmic assistance does not itself undermine commercial wisdom. The difficulty arises where opaque technological recommendations substitute for meaningful creditor evaluation or prevent affected stakeholders from identifying material errors. It proposes a framework of responsible algorithmic creditor control founded upon material-use disclosure, human-in-command oversight, auditability, model validation, identifiable responsibility and proportionate contestability. Artificial intelligence may augment creditor judgment, but it should not displace the human responsibility upon which creditor control and legitimate judicial restraint depend.

Keywords: Artificial Intelligence; Corporate Insolvency; Committee of Creditors; Commercial Wisdom; Algorithmic Decision Support; Creditor Control; Explainable AI; Human Oversight.

* PhD Research Scholar at National Law University Odisha, Cuttack, Odisha, India.

[1] Felix Steffek, *A Story of Two Holy Grails: How Artificial Intelligence Will Change the Design and Use of Corporate Insolvency Law*, U. CHI. L. REV. ONLINE (2024); Kayode Akintola, *A New Frontier? Exploring Artificial Intelligence in Corporate Insolvency*, 15 LAWS 79 (2026).

[2] Regulation (EU) 2024/1689 (Artificial Intelligence Act), arts. 9–15, annex III ¶ 5(b); Reserve Bank of India, RBI (Digital Lending) Directions, 2025, RBI/2025-26/36, DOR.STR.REC.19/21.07.001/2025-26 (May 8, 2025).

1 INTRODUCTION: INSOLVENCY GOVERNANCE ENTERS THE DIGITAL DECADE

Corporate insolvency law allocates decision making authority during periods of acute financial distress. Once a company can no longer meet its obligations in an orderly manner, the legal system determines who should control the future of the enterprise, how competing claims should be evaluated and whether restructuring or liquidation offers the more appropriate outcome. Under the Insolvency and Bankruptcy Code, 2016, these questions are answered principally through institutions governed by human judgment. Financial creditors exercise commercial authority through the Committee of Creditors, insolvency professionals administer the process and adjudicatory institutions supervise compliance with statutory requirements. The informational foundations of these decisions, however, are increasingly shaped by artificial intelligence, machine learning, predictive analytics and automated decision support systems.[3]

Algorithmic tools are already changing how credit risk is assessed, monitored and managed. Credit scoring systems, behavioral analytics, automated risk management platforms and early warning models enable financial institutions to process large volumes of transactional, operational and market information. The Reserve Bank of India's Digital Lending Directions, 2025 expressly regulate matters including creditworthiness assessment, borrower disclosure, data collection, privacy and technology standards within digital lending. These developments demonstrate that algorithmic systems already influence important aspects of the creditor-debtor relationship before insolvency begins. Their potential relevance does not end when financial distress develops. The same analytical capabilities may support claims administration, investigation, valuation, recovery estimation and assessment of restructuring alternatives.[4]

Emerging research confirms some use of artificial intelligence in corporate insolvency proceedings, including in tasks connected with investigations, asset management and creditor claims. Akintola's survey-based study reports that efficiency is a principal rationale for such use, while doubts regarding the accuracy of algorithmic outputs remain a central concern. Steffek separately examines how artificial intelligence may affect corporate insolvency law by improving prediction of legal outcomes and the design of contracts concerning financial distress. These contributions establish that artificial intelligence has practical and theoretical significance for insolvency law. They do not, however, fully address how material reliance upon algorithmic recommendations affects the legitimacy of creditor control or the judicial restraint traditionally associated with the Committee's commercial wisdom.[5]

This paper does not assume that Indian Committees of Creditors presently delegate votes or resolution decisions to autonomous systems. Nor does it claim that algorithmic systems have displaced human representatives within the statutory architecture of the IBC. Its concern is the legally significant continuum between existing digital risk assessment and the increasing technical capacity of algorithms to influence valuation, cash flow modelling, recovery analysis, viability assessment and comparison of resolution alternatives. At one end of this continuum, automated systems perform routine administrative tasks according to predetermined instructions. At the other, an algorithm may effectively determine an outcome without meaningful human evaluation. Between these extremes lies algorithmic decision support, where formal authority remains human but the assumptions, rankings and recommendations underlying the decision are materially shaped by technology.

---

[3] Insolvency and Bankruptcy Code, 2016, §§ 21, 24, 28, 30, 31.

[4] Reserve Bank of India, RBI (Digital Lending) Directions, 2025, RBI/2025-26/36, DOR.STR.REC.19/21.07.001/2025-26, chs. III–IV (May 8, 2025).

[5] Steffek, *A Story of Two Holy Grails*; Akintola, *A New Frontier?*.

This middle category creates a practical difficulty for creditor-controlled insolvency. The Supreme Court's commercial wisdom jurisprudence recognizes the primacy of the Committee's assessment of feasibility, viability and commercial outcomes, while confining adjudicatory review principally to the statutory boundaries imposed by the Code. Judicial restraint is justified not simply because a vote has been formally cast by creditors, but because financial creditors are presumed to possess the expertise, information and economic incentive necessary to make informed commercial assessments. Reliance upon technology does not by itself contradict that premise. Creditors already depend upon valuations, financial advisers, technical reports and internal risk systems. The difficulty arises when reliance becomes abdication: where creditor representatives accept an algorithmic recommendation without meaningful inquiry, cannot identify the material assumptions underlying it or are unable to depart from its conclusion.[6]

The difficulty becomes sharper because of opacity and unequal technological capacity. Imagine a creditor rejecting a resolution proposal because an internal model predicts weak recovery. The corporate debtor, operational creditors and even the resolution professional may know the result without knowing which assumptions or data limitations produced it. Historical or incomplete data may distort the recommendation, while automation bias may make a numerical output appear more objective than it really is. Large institutions may also have proprietary models and specialist teams that smaller creditors and other stakeholders cannot match. Technology can therefore reduce one information gap while creating another.[7]

Regulatory developments outside insolvency law offer useful, though limited, guidance. The European Union Artificial Intelligence Act adopts a risk-based framework and classifies specified AI systems used to evaluate the creditworthiness of natural persons or establish credit scores as high-risk. It also establishes requirements regarding risk management, data governance, documentation, transparency, human oversight, accuracy and robustness for high-risk systems. An algorithm used by a corporate creditor to evaluate an insolvency resolution plan does not automatically fall within that listed natural person creditworthiness category. The Regulation is therefore relevant here as a source of governance principles rather than as a rule directly governing Committee of Creditors voting. Similarly, the RBI Digital Lending Directions regulate lending activity rather than collective insolvency decision making. Their significance lies in the principle that regulated institutions retain responsibility for technology enabled financial activity and cannot transfer accountability entirely to technology vendors.[8]

Against this background, the primary question addressed by this paper is: when creditor decisions are materially influenced by algorithmic recommendations, under what conditions should those decisions remain protected by the judicial restraint traditionally associated with commercial wisdom? Two subsidiary questions follow. First, which forms of algorithmic involvement are sufficiently influential to require legal safeguards? Secondly, what standards of disclosure, oversight, auditability, validation, responsibility and contestability are required to preserve meaningful human commercial judgment?

The paper uses doctrinal analysis of the IBC and the Supreme Court's commercial wisdom jurisprudence, functional analysis of algorithmic involvement in insolvency related decision making and

---

[6] K. Sashidhar v. Indian Overseas Bank, (2019) 12 SCC 150; Comm. of Creditors of Essar Steel India Ltd. v. Satish Kumar Gupta, (2020) 8 SCC 531.

[7] Linda J. Skitka, Kathleen L. Mosier & Mark Burdick, *Does Automation Bias Decision-Making?*, 51 INT'L J. HUM.-COMPUT. STUD. 991–1006 (1999).

[8] Regulation (EU) 2024/1689 (Artificial Intelligence Act), arts. 9–15, annex III ¶ 5(b); RBI Digital Lending Directions, 2025, chs. III–IV.

comparative examination of emerging European and Indian digital regulation principles. Its contribution is narrower than a general account of artificial intelligence in corporate insolvency. It argues that the legitimacy of algorithmically informed creditor control depends upon whether legal authority remains connected to meaningful human evaluation and identifiable responsibility. The paper therefore proposes a framework of responsible algorithmic creditor control founded upon material-use disclosure, human-in-command oversight, auditability, model validation, functional allocation of responsibility and proportionate contestability. Artificial intelligence may strengthen the informational basis of creditor judgment, but it should not replace the human responsibility upon which creditor control and legitimate judicial restraint depend.

2 FROM CREDITOR CONTROL TO ALGORITHMICALLY INFORMED CREDITOR CONTROL

Corporate insolvency law does more than provide remedies for unpaid debt. It reallocates control over a distressed enterprise and determines which actors may decide whether the business should be restructured, sold or liquidated. Modern insolvency regimes commonly employ collective decision-making structures because individual enforcement may fragment enterprise value and produce outcomes that are inferior to coordinated resolution. The allocation of authority therefore lies at the centre of insolvency governance.

The Insolvency and Bankruptcy Code, 2016 adopts a creditor-in-control model for the corporate insolvency resolution process. Following commencement, the powers of the board of directors are suspended and the management of the corporate debtor is entrusted to the interim resolution professional and, subsequently, the resolution professional. Financial creditors constitute the Committee of Creditors, which functions as the principal commercial decision-making body within the resolution process. The Committee may take decisions on matters requiring its approval, evaluate resolution plans and determine whether the available proposals offer a feasible and viable basis for preserving or reorganising the corporate debtor.[9]

The statutory architecture distributes authority among three principal institutions. The Committee exercises commercial decision-making power. The resolution professional administers the process, collates and verifies claims, manages the corporate debtor as a going concern, prepares relevant information and facilitates the receipt and consideration of resolution plans. The National Company Law Tribunal exercises adjudicatory supervision, including examination of whether an approved plan complies with the requirements prescribed by the Code. This allocation separates commercial evaluation from process administration and legal supervision.[10]

The 2026 amendments retain this basic distribution while modifying parts of the resolution architecture. Among other reforms, the amended framework strengthens procedural timelines, introduces a creditor-initiated insolvency resolution process, requires the Committee to record reasons for approving resolution plans and revises aspects of plan approval and implementation. These developments strengthen the importance of documented creditor decision-making but do not replace the Committee as the central institution responsible for commercial evaluation. The current analysis should therefore be understood

[9] Insolvency and Bankruptcy Code, 2016, §§ 17, 21, 24, 28, 30.
[10] *Id.* §§ 18, 25, 30, 31.

as applying to creditor decisions under the amended framework as well as to the conventional corporate insolvency resolution process.[11]

Judicial interpretation has reinforced the centrality of CoC commercial judgment. In *K. Sashidhar v. Indian Overseas Bank*, the Supreme Court rejected substantive judicial substitution of the commercial determination made by the Committee. In *Committee of Creditors of Essar Steel India Ltd. v. Satish Kumar Gupta*, the Court reaffirmed the primacy of the CoC's assessment of feasibility, viability and distribution, while recognising the statutory limits within which that assessment must operate. The tribunal's task is not to formulate the commercial outcome that it considers preferable, but to examine whether the decision and the resulting plan satisfy the requirements imposed by the Code. This restraint is reinforced by *Swiss Ribbons*, *Maharashtra Seamless* and *Kalpraj Dharamshi*, which place commercial evaluation with creditors and limit judicial re-assessment of plan value or comparative merits.[12]

This approach is usually described as "commercial wisdom". That expression should not be understood as an assertion that creditor decisions are infallible or immune from every legal constraint. It describes the location of commercial authority within the statutory process and the correspondingly restrained role of adjudicatory institutions in reviewing commercial merits. The Committee remains subject to statutory voting requirements, plan compliance provisions, procedural obligations and the limited review expressly contemplated by the Code.

Creditor control rests on several connected assumptions. Financial exposure gives creditors an incentive to preserve and maximise value. Financial institutions may possess experience in credit evaluation, restructuring and assessment of commercial risk. Collective decision-making can reduce coordination problems among creditors. Commercial questions concerning viability, recoveries and restructuring may also be less suitable for determination through broad judicial discretion. These propositions support judicial restraint only where the Committee's decision represents a genuine exercise of informed commercial judgment.

It is important to distinguish decision authority and decision influence. Decision authority concerns the legal power to vote, approve a resolution plan or determine whether liquidation should be pursued. Decision influence concerns the information, assumptions, professional advice, valuation reports, risk assessments and analytical tools that shape how that legal power is exercised. A creditor does not surrender its authority merely because it relies upon advice or technology. Creditors have always depended upon financial experts, valuers, legal advisers and internal risk systems. The relevant inquiry is whether those inputs assist the exercise of commercial judgment or effectively substitute for it.

Algorithmic technologies intensify this distinction. Credit risk models, predictive analytics, machine learning applications and automated valuation tools may identify patterns, estimate outcomes and recommend preferred alternatives. Even where an algorithm possesses no legal power to vote, its output may determine which risks are emphasised, which scenarios are considered credible and which resolution proposal appears commercially preferable. Formal authority remains with the creditor, but the informational architecture underlying the decision may increasingly be technological.

This does not suggest that algorithmic influence necessarily diminishes creditor control. A well-designed system may improve the quality, speed and consistency of information available to a creditor

[11] Insolvency and Bankruptcy Code, 2016, §§ 17, 21, 24, 28, 30; Insolvency and Bankruptcy Code (Amendment) Act, 2026.

[12] *K. Sashidhar*; *Essar Steel*; Swiss Ribbons Pvt. Ltd. v. Union of India, (2019) 4 SCC 17; Maharashtra Seamless Ltd. v. Padmanabhan Venkatesh, (2020) 11 SCC 467; Kalpraj Dharamshi v. Kotak Inv. Advisors Ltd., (2021) 10 SCC 401.

representative. The concern arises where technological influence becomes sufficiently dominant that meaningful human evaluation cannot be demonstrated. If representatives do not understand the system's function, cannot examine material assumptions, fail to consider relevant limitations or treat the recommendation as conclusive, decision support may become an abdication of judgment. The legal problem is therefore not the presence of an algorithm but the absence of meaningful human commercial assessment.

Algorithmic systems may enter before insolvency through credit assessment, transaction monitoring, covenant review and risk management. They may identify deterioration and influence forbearance, enforcement or restructuring engagement, but remain informational inputs rather than legal findings of default or viability.[13]

They may also organise contractual records, payment histories and evidentiary material relevant to commencement. This role is primarily administrative; responsibility for commencing proceedings and proving statutory conditions remains with the actors identified by the Code.

Evidence of current use is strongest in process support. Artificial intelligence may assist insolvency professionals and advisers with claims administration, document management, anomaly detection, asset tracing and transaction review. These tools expand investigative capacity but support rather than replace legal and professional determination; their influence on creditors is often indirect through reports prepared by professionals, valuers and advisers.[14]

During plan evaluation, computational tools may estimate recovery ranges, test cash flow assumptions and compare restructuring alternatives. Current evidence does not establish routine autonomous ranking of Indian resolution plans, so these applications should be treated as emerging decision support. Their design may nevertheless privilege immediate recovery over resilience, employment continuity or longer-term viability.

Three degrees of involvement should be distinguished. Automated execution performs routine tasks under predetermined instructions. Algorithmic decision support produces predictions, scores or recommendations while authority remains human. Autonomous decision-making effectively determines the outcome without meaningful consideration or practical override. The classification is functional: a nominally advisory tool may operate autonomously if its outputs are routinely accepted without examination.

Responsibility must remain with the actor exercising legal or professional authority. Creditors remain accountable for votes, professionals for statutory functions, advisers for their analyses and vendors for matters within system design and performance obligations. Meaningful override requires sufficient understanding to question outputs, examine assumptions, recognise limitations and depart from a recommendation. Algorithms become problematic only when material and opaque influence makes independent commercial judgment unidentifiable.

3 THE PROMISE OF ALGORITHMIC INSOLVENCY GOVERNANCE

Debate about AI often begins with risk, but this should not obscure why algorithmic technologies attract attention. Insolvency takes place under information asymmetry, time pressure and uncertainty. Stakeholders must evaluate complex data while making decisions affecting enterprise survival and loss allocation.

[13] RBI Digital Lending Directions, 2025, chs. III–IV.
[14] Akintola, *A New Frontier?*.

A principal advantage is the capacity to detect financial deterioration earlier than traditional monitoring. Predictive analytics may identify missed obligations, liquidity pressures, deteriorating cash flow and covenant breaches. Systems can continuously examine transactional and operational information rather than await periodic reviews. In theory, they permit intervention before distress develops into full-scale insolvency.[15]

This capacity is significant because delay destroys value. As distress deepens, assets deteriorate, customers leave, financing contracts and restructuring prospects weaken. The Bankruptcy Law Reforms Committee identified delay as a serious weakness of the prior insolvency system and emphasised early intervention in preserving enterprise value. Algorithms capable of identifying deterioration may therefore support one of the central objectives of modern insolvency law.[16]

Their value does not lie in declaring insolvency automatically. Insolvency remains a legal determination requiring human judgment and institutional oversight. Their principal benefit is informational: alerting creditors, management and professionals to developing risks while corrective action remains feasible. They are most valuable when they facilitate human intervention rather than replace it.

Information asymmetry is a persistent problem. Management generally possesses superior knowledge of operations, liquidity, obligations and prospects, while creditors and other stakeholders make decisions on incomplete information. Creditors may struggle to evaluate a restructuring plan, determine asset value or assess the credibility of management projections.

The creditor-in-control model presupposes informed commercial judgment, yet Committee members may need to review extensive documentation, operational records and financial information. The challenge is not merely obtaining information but identifying what matters. Algorithmic systems may enhance capacity by identifying patterns, anomalies and relevant information across large datasets. They may supplement commercial judgment rather than replace it.

Their value is evident in valuation and recovery analysis. Creditors compare future outcomes, assess whether restructuring offers a better return than liquidation and examine projected cash flows. Algorithmic tools may supplement financial reports and expert opinions through analysis of historical performance, sectoral trends and scenarios. Shared analytical tools might also help smaller creditors understand plan assumptions and recovery ranges.

Further, suppose the Committee is comparing two plans: one offers higher immediate payment, while the other keeps more of the business and workforce intact but pays creditors over a longer period. An analytical model may test cash flows, sector trends and likely recovery under each option. Used carefully, it can sharpen the discussion and help smaller creditors understand the range of outcomes. It should not, however, be treated as a neutral answer to a question that still requires commercial judgment.[17]

AI also promises consistency and administrative efficiency. Insolvency professionals, creditors and tribunals process large quantities of records under time constraints. Technologies can classify documents, reconcile claims, identify unusual transactions and extract relevant provisions. Artificial intelligence may reduce administrative burdens and allow human actors to focus on legal and commercial judgment.[18]

---

[15] *Id.*

[16] BANKRUPTCY LAW REFORMS COMM., THE REPORT OF THE BANKRUPTCY LAW REFORMS COMMITTEE, VOLUME I: RATIONALE AND DESIGN chs. 3–4 (Nov. 2015).

[17] Steffek, *A Story of Two Holy Grails*; Akintola, *A New Frontier?*.

[18] Akintola, *A New Frontier?*.

Consistency is another attraction. Human decision-making is vulnerable to fatigue, overload and inconsistency. Algorithms can apply the same analytical framework repeatedly across claims or records. Such consistency may improve reliability in routine processing and risk analysis.

Yet consistency is not correctness. A system may repeatedly produce the same output while relying upon flawed data or unsuitable assumptions. Algorithmic errors may be replicated at a greater scale than human errors. Efficiency cannot justify abandonment of scrutiny regarding accuracy, transparency and accountability. Insolvency law does not exist solely to process cases rapidly; it balances value preservation, fairness, participation and procedural integrity. Speed is valuable only where the resulting process remains legitimate and trustworthy.[19]

4 LEGAL RISKS OF ALGORITHMICALLY INFLUENCED CREDITOR DECISIONS

The same tools that improve efficiency may also create opacity, bias and accountability deficits. Insolvency systems derive legitimacy from the ability of stakeholders to understand, challenge and evaluate decisions that affect their rights. The practical issue is not whether technology should be used, but under what conditions algorithmic influence remains compatible with transparency, procedural fairness and accountability.

Opacity is the most frequently discussed challenge. Insolvency law assumes that significant decisions can be justified through intelligible reasons. Algorithms complicate this assumption. A creditor may reject a plan because an internal model classifies it as risky, but the model may rely upon thousands of variables and generate conclusions through mechanisms difficult to explain even to its users.

Different forms of explainability should be distinguished. Full disclosure of models, training methods and source code may be impractical and commercially sensitive. A more realistic approach distinguishes model transparency from decision transparency. Insolvency law may require disclosure of a system's function, the principal factors materially influencing its recommendation and relevant limitations, without requiring publication of complete technical architecture.[20]

Explainability matters because stakeholders need confidence in the process. If a proposal is rejected because a model predicts insufficient recovery, saying only that the plan received a low score is not enough. The affected parties should at least understand whether the concern arose from projected cash flow, valuation assumptions, sector risk or implementation uncertainty. Without that basic explanation, there is little room to identify an error or correct a false assumption.[21]

The issue is also institutional. Large financial institutions may have advanced analytics, proprietary datasets and technical expertise unavailable to smaller creditors. Although all creditors may formally participate, some may possess much greater capacity to shape outcomes. Bias may arise from deficient data, but it also operates through unequal distribution of technological capacity.

---

[19] EUR. BANKING AUTH., AI ACT: IMPLICATIONS FOR THE EU BANKING AND PAYMENTS SECTOR (Nov. 21, 2025).

[20] EUR. BANKING AUTH., AI ACT: IMPLICATIONS FOR THE EU BANKING AND PAYMENTS SECTOR; Artificial Intelligence Act arts. 13–14; Finale Doshi-Velez & Been Kim, *Towards a Rigorous Science of Interpretable Machine Learning* (arXiv:1702.08608, 2017); Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIV. L. 76–99 (2017).

[21] Akintola, *A New Frontier?*; HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, ETHICS GUIDELINES FOR TRUSTWORTHY AI (Eur. Comm'n 2019).

Traditional law allocates responsibility to identifiable actors: creditors vote, professionals administer and courts decide. Algorithmic support complicates responsibility because developers, vendors, lenders and Committee representatives may all contribute to an output.

If a model proves inaccurate, the developer may blame data, the creditor may call it advisory and the user may characterise reliance as reasonable. This diffusion risks leaving no actor accountable. Insolvency law should reject an "algorithm made the decision" defence. Where creditors exercise legal authority, they remain responsible for decisions adopted on the basis of algorithmic recommendations.[22]

Responsibility should be allocated according to function. Developers may be responsible for design and known limitations; institutions for deployment, validation and data governance; individual decision-makers for the final adoption of recommendations. Algorithms influence human decisions without replacing human responsibility.

Opacity, bias and accountability converge upon procedural fairness. Insolvency decisions affect creditors, debtors, employees and investors. Stakeholders should be able to challenge materially inaccurate information or demonstrably flawed assumptions.

This does not require full natural-justice hearings regarding every algorithm. Contestability should be proportional to the significance of the system's role. Stronger influence over substantive outcomes justifies stronger disclosure, correction and review. The objective is to preserve confidence without turning every model assessment into separate litigation.[23]

The final challenge concerns commercial wisdom. Judicial restraint is justified because creditors are presumed to possess expertise and incentives to evaluate outcomes. Yet if representatives routinely accept algorithmic recommendations without independent assessment, judicial restraint may protect a technological output rather than commercial judgment.

Automation bias is the tendency to place disproportionate trust in algorithmic outputs, particularly in complex and uncertain environments. Insolvency has precisely those characteristics. This creates a doctrinal paradox: limited judicial review is justified by informed creditor judgment, but opacity may prevent verification that such judgment occurred. The challenge is therefore to preserve meaningful human responsibility within a technologically influenced process.[24]

5 Comparative Regulatory Signals for Insolvency Law

Algorithmic challenges are not unique to insolvency. Regulators across sectors are addressing transparency, accountability, bias and human oversight. Although most AI regulation was developed outside insolvency, it provides guidance for future creditor decision-making.

Risk-based frameworks distinguish AI systems according to the consequences associated with their use. The European Union Artificial Intelligence Act establishes differing obligations and enhanced duties for high-risk systems. AI used to evaluate the creditworthiness of natural persons or establish credit scores is listed as a high-risk use.[25]

---

[22] Artificial Intelligence Act arts. 16, 26; RBI Digital Lending Directions, 2025, chs. III–IV.

[23] RBI Digital Lending Directions, 2025, chs. III–IV.

[24] Skitka, Mosier & Burdick, *Does Automation Bias Decision-Making?*; Raja Parasuraman & Victor Riley, *Humans and Automation: Use, Misuse, Disuse and Abuse*, 39 HUM. FACTORS 230–53 (1997); *K. Sashidhar*; *Essar Steel.*

[25] Artificial Intelligence Act annex III ¶ 5(b).

The framework emphasises risk management, data governance, documentation, record keeping, transparency, human oversight, accuracy and robustness. These principles are relevant to insolvency even though plan ranking or creditor voting tools may not fall within the same listed category. Comparable concerns arise whenever algorithmic outputs materially affect significant economic interests. The OECD AI Principles similarly emphasise transparency and explainability, robustness and accountability.[26]

The risk-based model also demonstrates that not all AI uses require identical regulation. Automated document classification may warrant lighter safeguards than tools used to rank resolution plans or estimate recoveries. Regulation should be proportionate to function and influence.

Financial sector regulation provides a second reference point. The Reserve Bank of India's Digital Lending Directions, 2025 address creditworthiness assessment, borrower disclosures, data collection and use, privacy, technology standards and responsibility of regulated entities. The Directions reflect a principle that technological innovation must remain connected to transparency, customer protection and institutional responsibility.[27]

This is relevant to insolvency because algorithmic responsibility can become diffuse. Financial regulation maintains accountability at the level of the regulated entity rather than allowing it to shift entirely to vendors. The same principle can govern insolvency: creditor institutions remain responsible for technological systems used to inform their legal decisions.

Digital lending regulation does not fully address collective insolvency. It focuses primarily upon origination and servicing, while insolvency involves multiple stakeholders with competing claims. It nevertheless demonstrates a regulatory convergence around transparency, accountability, human oversight, data governance and proportionality.

Existing insolvency law contains a variety of mechanisms capable of reviewing decisions, supervising insolvency professionals and ensuring procedural compliance. Commercial wisdom jurisprudence, statutory duties imposed upon insolvency professionals and existing principles of administrative fairness all perform important governance functions within the insolvency process. Even so, these mechanisms were largely developed in an environment where material decision-making was assumed to be explainable through identifiable human reasoning. They do not adequately address circumstances in which influential recommendations are generated by opaque algorithmic systems, where material assumptions may be inaccessible to affected stakeholders and where responsibility becomes fragmented among creditors, advisers and technology providers. Existing legal doctrine therefore provides only a partial response to the governance challenges created by algorithmically informed creditor control.

Any regulatory framework governing algorithmically informed creditor control must remain sensitive to commercial realities. Excessively burdensome disclosure obligations may increase compliance costs, discourage innovation and reduce the ability of creditors to respond quickly to deteriorating financial circumstances. Broader transparency requirements may also raise legitimate concerns regarding confidential business information, proprietary analytical models and commercially sensitive restructuring strategies. The objective of regulation should therefore not be the elimination of technological risk through exhaustive disclosure. Rather, it should be the creation of proportionate safeguards that preserve

[26] Artificial Intelligence Act arts. 9–15; ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, RECOMMENDATION OF THE COUNCIL ON ARTIFICIAL INTELLIGENCE, OECD/LEGAL/0449 (2019).
[27] RBI Digital Lending Directions, 2025, chs. III–IV.

accountability and procedural legitimacy while allowing creditors to retain the flexibility necessary for effective commercial decision-making.

6 A Regulatory Framework for Responsible Algorithmic Creditor Control

Neither unrestricted deployment nor prohibition is a sensible answer. A document classification tool does not raise the same concerns as a model that ranks resolution plans. The safeguards should therefore match the function of the system and the weight given to its output. Six connected principles provide a workable starting point. These principles also reflect convergence across the AI Act, OECD AI Principles and the Ethics Guidelines for Trustworthy AI.[28]

Material-use disclosure. Where an algorithm materially influences valuation, recovery assessment, plan comparison or a creditor recommendation, the nature and extent of that use should be disclosed in a proportionate form. Disclosure need not reveal source code or commercially sensitive technical architecture, but should identify the system's function, the principal categories of information considered and any material limitations affecting reliance.

Human-in-command oversight. Formal human approval is insufficient. The responsible representative must understand the system's purpose, interrogate outputs, evaluate material assumptions, recognise limitations and reject recommendations where appropriate. Creditors cannot avoid responsibility by attributing the conclusion to a model; meaningful engagement preserves the basis of commercial wisdom.[29]

The same standard requires substantive engagement rather than passive acceptance of technological output.[30]

Auditability and record preservation. Materially influential systems should preserve the model version, intended purpose, input categories, output and the decision-maker's response. Retrospective examination supports accountability without continuous monitoring or public disclosure of source code.[31]

Data quality and validation. A model must be validated for the insolvency function in which it is used. Systems designed for loan origination may be unsuitable for restructuring evaluation; validation should examine assumptions, performance, suitability, drift and context.[32]

Allocation of responsibility. Outsourcing technology must not outsource accountability. Creditors remain responsible for votes; resolution professionals for statutory duties and process information; advisers for analyses and valuations; and vendors for design, documentation and system performance within their obligations.[33]

Resolution professionals should secure process integrity and disclosure compliance without becoming guarantors of every creditor's proprietary model.

Proportionate contestability. Stakeholders should be able to challenge false data, undisclosed reliance or absence of meaningful human consideration. Review should correspond to the system's influence and

---

[28] Eur. Banking Auth., AI Act: Implications for the EU Banking and Payments Sector; Organisation for Economic Co-operation and Development, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449 (2019); High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI (Eur. Comm'n 2019); Artificial Intelligence Act arts. 9–15.

[29] Artificial Intelligence Act art. 14; Steffek, *A Story of Two Holy Grails*; John D. Lee & Katrina A. See, *Trust in Automation: Designing for Appropriate Reliance*, 46 Hum. Factors 50–80 (2004).

[30] Wachter, Mittelstadt & Floridi, *Why a Right to Explanation*.

[31] Artificial Intelligence Act arts. 11–12.

[32] Akintola, *A New Frontier?*; Artificial Intelligence Act arts. 9–10, 15.

[33] RBI Digital Lending Directions, 2025; Artificial Intelligence Act arts. 16, 26.

focus on legality and procedural integrity rather than substitution of judicial preferences for commercial judgment.[34]

Contestability does not require disclosure of confidential source code. Relevant inputs, material assumptions, methodological explanations and the rationale for a recommendation may permit meaningful scrutiny while protecting legitimate intellectual property.[35]

Consistently with existing doctrine, tribunals need not become technical regulators of AI; they should ensure that materially influential use satisfies minimum standards of transparency, accountability and fairness.[36]

7 CONCLUSION

The digital transformation of corporate finance is altering the informational foundations of insolvency. Artificial intelligence, predictive analytics and algorithmic decision-support are moving from credit origination and risk management into distress monitoring, claims analysis, valuation, recovery estimation and resolution planning. The future of creditor control cannot therefore be analysed solely through concepts developed for exclusively human decision-makers.[37]

This paper has argued that algorithmic insolvency governance presents opportunities and risks. AI may support earlier identification of distress, reduce information asymmetry and improve administrative efficiency. Properly deployed technologies may improve the quality and timeliness of information available to creditors and insolvency professionals.[38]

At the same time, opacity may undermine explainability; biased or incomplete data may distort outcomes; technological asymmetry may concentrate informational power; and diffusion of responsibility may frustrate accountability. Most importantly, dependence upon algorithmic recommendations challenges commercial wisdom. Judicial restraint assumes genuine informed judgment. If judgment is displaced by opaque technological outputs, the normative basis of judicial restraint is weakened.

The analysis rejects two extremes. Technological resistance would sacrifice useful analytical capacity. Technological determinism would treat algorithmic outputs as inherently objective and disregard transparency, accountability and fairness. Instead, the paper proposes responsible algorithmic creditor control. Algorithms should function as decision support mechanisms rather than autonomous decision-makers. Human actors remain responsible, while technology operates within safeguards of material-use disclosure, meaningful human oversight, auditability, data validation, clear responsibility and proportionate contestability.[39]

The central challenge is not the use of artificial intelligence within insolvency processes, but the changing relationship between legal authority and technological influence. Under the Insolvency and Bankruptcy Code, commercial authority remains vested in creditors acting through the Committee of Creditors. Algorithmic systems may assist those creditors by improving information processing, modelling potential outcomes and identifying risks that may not be immediately apparent through traditional

---

[34] Artificial Intelligence Act arts. 13–14; RBI Digital Lending Directions, 2025, chs. III–IV.

[35] Wachter, Mittelstadt & Floridi, *Why a Right to Explanation*.

[36] Insolvency and Bankruptcy Code, 2016, §§ 30–31; *Essar Steel*.

[37] Steffek, *A Story of Two Holy Grails*; Akintola, *A New Frontier?*.

[38] Akintola, *A New Frontier?*; BANKRUPTCY LAW REFORMS COMMITTEE REPORT chs. 3–4.

[39] Artificial Intelligence Act arts. 9–15; EUR. BANKING AUTH., AI ACT: IMPLICATIONS FOR THE EU BANKING AND PAYMENTS SECTOR.

analysis. Yet the existence of formal authority is not, by itself, sufficient to justify judicial restraint. The legitimacy of creditor control depends upon the continued exercise of meaningful human commercial judgment. The stronger the algorithmic influence upon a creditor decision, the stronger the justification for corresponding obligations of transparency, accountability and human oversight. Where technologically generated recommendations become effectively determinative and the basis of those recommendations cannot be identified, tested or challenged, the rationale underlying commercial wisdom becomes progressively weaker.

The future of creditor-controlled insolvency will not be resolved by choosing between people and artificial intelligence. It will depend on how well legal institutions govern their interaction. Algorithms can help creditors process information, model outcomes and identify risks that may otherwise be missed. They should not, however, displace the identifiable human responsibility on which creditor authority rests. For a supplier waiting for payment, an employee concerned about job security or a debtor trying to save a viable business, the reasons behind an insolvency decision matter. The proposed safeguards, like material-use disclosure, human-in-command oversight, auditability, validation, clear responsibility and proportionate contestability, seek to preserve that human centre. AI should remain a tool of insolvency governance, not its substitute.

* * *

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.18

CHAPTER 18

Cyber Risk Insurance for Fintech and Digital Payments: Regulatory Challenges and Emerging Coverage Gaps in India

SOWMYA. H.A*

Fintech platforms have shown an explosive growth with the rapid adoption of digital payments. However, the same has exposed especially Indian fintech and digital payment platforms to cyber risks, i.e., incidents such as phishing, payment fraud, ransomware, data breaches, identity theft, system disruption, etc. Moreover, companies and businesses have become more vulnerable to third-party incidents due to increased exposure to the cyber domain. Cyber risk insurance is one of the mechanisms to mitigate risks by transferring the financial losses to the insurer.

However, it is not clear whether the existing legal provisions and insurance policies provide sufficient coverage for fintech and digital payment platforms. This paper examines the question, to what extent does the existing legal framework facilitate cyber risk insurance, and how effective is it in managing the emerging cyber risks in India?

The paper adopts doctrinal and analytical approaches to study the relevant sections of the legislation, regulatory directives, and judicial developments. The paper also examines the regulatory frameworks of payment and cyber risks by the Reserve Bank of India (RBI) and the Insurance Regulatory and Development Authority of India (IRDAI) and addresses the growing concerns in fintech ecosystems.

International practices are also examined based on existing regulatory frameworks for cyber and digital payment risks and cyber risk insurance, and their relevance to India is identified.

The paper concludes that the regulatory framework in India for cyber risk and digital payments is evolving to identify gaps in risk transfer, thereby highlighting the need for a progressive approach in cyber risk insurance. The paper identifies relevant regulatory actions and reforms to address these emerging gaps.

Keywords: Cyber Risk Insurance, Cybersecurity Regulation, Data Breach, Digital Payments, Fintech.

1 INTRODUCTION

The financial industry in India is evolving rapidly, with a digital revolution in the delivery of banking, payment and other financial services. The emergence of banking services, electronic wallets, payment aggregators, mobile applications, internet banking, Unified Payments Interface (UPI), pre-paid instruments and other technology-driven financial services has led to the formation of an integrated financial ecosystem where most financial transactions can be completed in an almost instant manner. The new financial services have boosted financial inclusion and convenience but also pose new risks. The bigger the number of participants in a digital financial transaction, the greater the impact of a cyber incident. A

* PhD Scholar at KLE Law College (A Constituent College of KLE Technological University), Bengaluru, Karnataka, India.

disruption in a technology platform or service provider can have a knock-on impact on merchants, customers, financial institutions and other stakeholders in the payment ecosystem.[1]

The rapid expansion of digital payments has altered the profile of financial risk. Whilst traditional financial risks such as credit risk, liquidity risk and market risk remain and continue to evolve, this is increasingly exacerbated by cyber risks that threaten the confidentiality, integrity and availability of digital financial infrastructure. A successful cyberattack could result in direct financial loss but also involve regulatory sanctions, litigation costs, reputational damage, business disruption and erosion of consumer confidence. Moreover, because digital financial transactions are performed via networked platforms, including banks, payment systems and fintechs and technology and other service providers, the damages could extend beyond the financial institutions directly affected.[2]

Accordingly, cyber risk insurance has become increasingly important as a vehicle for transferring certain of the financial consequences of cyber incidents. However, insurance cannot replace cyber control measures and is rather seen as a line of defence providing financial protection when an insured event occurs. Therefore, the relationship between cyber regulation and cyber insurance is a key point of consideration in the fintech environment. In this context, RBI has been significantly tightening its oversight of cyber risk at various levels of financial institutions and payment system participants, and IRDAI has been creating policy direction for cyber insurance products.[3]

Cyber risk insurance is a relatively new risk transfer mechanism for covering losses resulting from cyber events. The effectiveness of cyber risk insurance in the Indian context depends critically upon coordination between the insurance and cybersecurity regulatory regimes. Cybersecurity regulations form the basis for establishing the controls expected of an organisation, whereas insurance contracts form the basis for determining everything that constitutes coverage and the extent of indemnification. Disputes arising between insurers and insureds over the interpretation and application of policy exclusions and warranties, disclosure obligations, causation and the classification of losses are amongst the challenges faced in the application of cyber insurance.[4]

1.1 Growth of Fintech and Digital Payments in India

India's fintech ecosystem has grown rapidly recently due to widespread adoption of digital payment technologies and technology-driven financial service providers. The fintech ecosystem includes banks, non-bank Payment System Operators (PSOs), payment aggregators, payment gateways, fintech platforms, technology providers and other players. However, fintech operators perform a diverse range of regulatory functions and hence, different legal obligations may apply to different categories of operator, depending upon which regulatory obligations are subsumed by a particular entity.[5]

---

[1] Payment and Settlement Systems Act, 2007, §§ 3–4; Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators (July 30, 2024).

[2] Reserve Bank of India, Master Direction on Information Technology Governance, Risk, Controls and Assurance Practices (Nov. 7, 2023); Reserve Bank of India, Master Direction on Digital Payment Security Controls (Feb. 18, 2021).

[3] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, IRDAI/NL/CIR/MISC/242/09/2021 (Sept. 8, 2021); Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

[4] Insurance Regulatory and Development Authority of India, Guidelines on Information and Cyber Security for Insurers and Insurance Intermediaries (2022); Reserve Bank of India, Master Direction on Information Technology Governance, Risk, Controls and Assurance Practices, *supra* note 2.

[5] Payment and Settlement Systems Act, 2007; Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

The RBI's regulatory approach recognises this diversity. The Master Directions on Cyber Resilience and Digital Payment Security Controls for non-bank PSOs issued in 2024 establish requirements concerning governance, risk assessment, security controls, vendor management, data security, incident response, business continuity and digital payment security. The Directions also recognise that PSOs may depend upon payment gateways, technology vendors and other third-party service providers and therefore require them to identify and manage cyber and technology risks arising from such relationships.[6]

This regulatory structure is significant for cyber insurance because the risk profile of a fintech organisation cannot be assessed solely by examining its internal information technology infrastructure. A fintech enterprise may depend upon cloud service providers, software vendors, payment processors, application developers, telecommunications providers and other participants in the digital payment chain. Consequently, cyber insurance underwriting increasingly needs to examine the broader technology ecosystem rather than treating cyber risk as an isolated organisational risk.[7]

High-frequency and high-severity digital threats including artificial intelligence-enabled attacks, sophisticated phishing campaigns, ransomware and supply-chain compromises present rapidly evolving risks for financial technology companies. Some of these risks may affect multiple financial institutions simultaneously and may therefore create correlated losses. The changing cyber risk environment calls for specialised underwriting practices capable of assessing not only the technical security of an individual insured but also its exposure to interconnected technology infrastructure and systemic cyber events.[8]

1.2 Cyber Risk and the Need for Insurance

The coverage available under cyber liability insurance can vary depending on the specific policy. First-party coverage can include costs incurred in incident response, forensic investigations, data restoration, business interruption and certain types of cyber extortion. Third-party coverage can provide protection against claims by customers, business partners and other affected parties. Some policies may also address regulatory investigation costs and payment fraud, subject to defined conditions, exclusions and definitions.[9]

The existence of a cyber insurance policy does not however guarantee coverage for every cyber-related loss. Traditional exclusions relating to fraud, contractual liability, war, infrastructure failure or intentional acts may interact with cyber-specific provisions in complex ways. An employee-led payment fraud involving social engineering may therefore raise questions concerning whether the loss should be classified as a cyber event, employee dishonesty, financial crime or an unauthorised electronic transaction.[10]

This classification problem is particularly significant in fintech because a single incident may simultaneously contain technological, financial and human elements. For example, a fraudulent payment may begin with phishing, involve the disclosure of authentication credentials, result in an authorised transfer and ultimately cause a financial loss to a customer or financial institution. The question of which insurance

---

[6] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

[7] *Id.*

[8] *Id.*; Reserve Bank of India, Master Direction on Information Technology Governance, Risk, Controls and Assurance Practices, *supra* note 2.

[9] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

[10] *Id.*; *see also* Insurance Regulatory and Development Authority of India, Guidelines on Information and Cyber Security for Insurers and Insurance Intermediaries, *supra* note 4.

policy responds to the loss may therefore depend upon the precise wording of the policy rather than merely the factual description of the cyber incident.

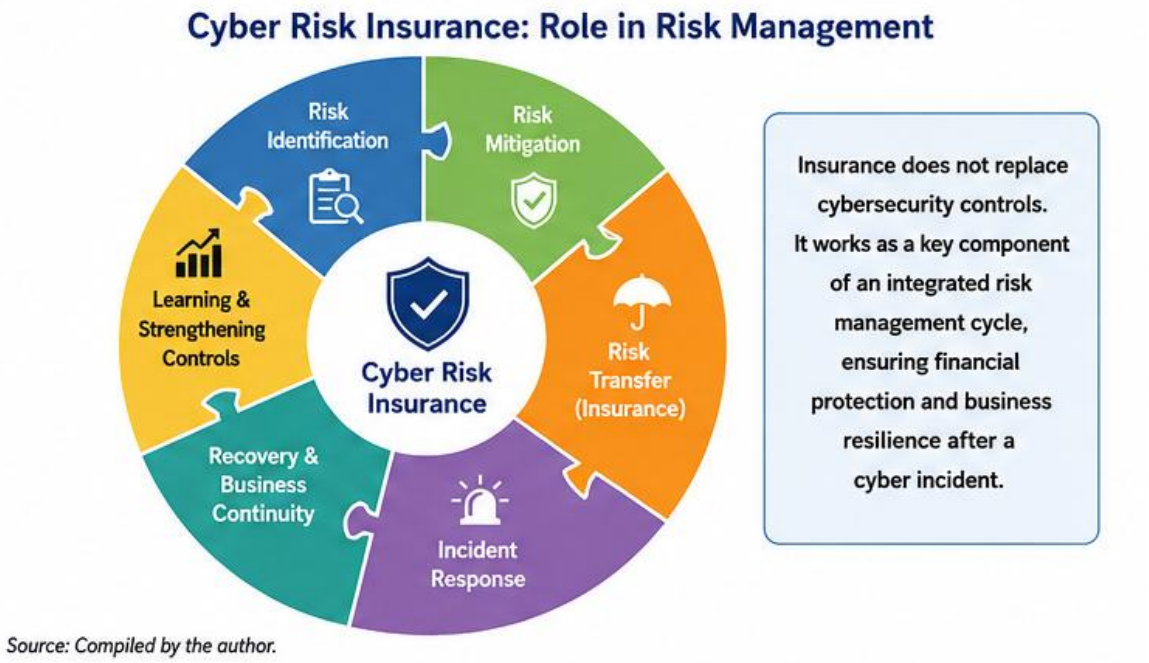

FIGURE 1. *Role of Cyber Risk Insurance in Risk Management*

2 CYBER RISKS AFFECTING FINTECH AND DIGITAL PAYMENTS

2.1 Payment Fraud and Social Engineering

Payment fraud presents significant cyber risks to digital financial services. Payment fraud incidents typically exploit social engineering tactics to coerce customers or employees into carrying out fraudulent transactions. Examples include phishing, vishing, business email compromise, impersonation and fraudulent instructions.[11]

Policies may provide limited coverage for these risks in that the loss situation resulting from a cybercrime attack, for instance, may be covered while loss resulting from a voluntary transfer of funds may be excluded. A more substantial provision is a social engineering coverage which clearly specifies exclusions that revolve around authentication, internal control and incident reporting requirements. These distinctions underline the importance of clear drafting in ensuring the payments and technology risks are understood.[12]

Under the RBI's regulatory framework for payment fraud prevention, payment fraud is increasingly recognized as a part of operational and cyber resilience that is worth monitoring and preventing. For example, the 2024 Directions require fraud monitoring and transaction monitoring in financial

[11] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.
[12] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

institutions in addition to other regulations that recognise payment frauds as one of the many ways to identify unauthorised or fraudulent digital transactions.[13]

Cyber insurance should complement these regulatory controls. Insurers may rely upon exclusions and warranties where an insured fails to maintain required safeguards. Nevertheless there is a need to distinguish intentional or reckless non-compliance from an isolated failure that occurs despite an otherwise reasonable cybersecurity programme. The legal consequences of a cybersecurity deficiency should ideally bear a rational relationship to the loss for which indemnification is claimed.[14]

2.2 Data Breaches and Identity Theft

Fintech organisations process substantial quantities of personal and financial information. Customer names, identification details, account information, transaction records, authentication information and other sensitive data may be processed across several technological environments. A data breach can therefore produce consequences extending beyond the immediate cost of repairing affected systems.[15]

The Information Technology Act, 2000 provides an important statutory foundation for addressing unauthorised access, damage to computer systems and computer-related offences. The Digital Personal Data Protection Act, 2023 subsequently established a dedicated legislative framework concerning the processing and protection of digital personal data. For financial technology organisations the interaction between cybersecurity obligations and data protection obligations is therefore increasingly important.[16]

For insurers, data breach risk may include forensic investigation costs, legal costs, notification costs, crisis management, remediation costs and third party claims. Insurance policy wording may also include provisions requiring certain security measures, incident notification and cooperation with the insurer. Such provisions are likely to be especially relevant where the insured is already subject to statutory data protection and incident reporting obligations.[17]

The statutory privacy obligations and the insurance policy wording may, however, be based on different concepts. A statutory reporting obligation may arise where personal data has been lost, even if the insured has not itself incurred an equivalent monetary loss. On the other hand, the insured could suffer lengthy interruption as a result of a cyber incident that does not involve loss or unauthorized access to personal data. Cyber insurance may thus be rendered ineffective if all losses are treated as data breach losses.[18]

2.3 Ransomware and Business Interruption

Ransomware is a growing threat for highly digitised organisations. Ransomware may block an organisation's access to its systems, affect payment processing and prevent customers from requesting financial services. Thus, even if personal data has not been irretrievably lost, significant losses can still occur.[19]

---

[13] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

[14] *Id.*; Reserve Bank of India, Master Direction on Digital Payment Security Controls, *supra* note 2.

[15] Digital Personal Data Protection Act, 2023; Information Technology Act, 2000.

[16] Information Technology Act, 2000; Digital Personal Data Protection Act, 2023.

[17] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

[18] *Id.*

[19] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

CERT-In's directions requiring reporting of cyber incidents are especially relevant in this context. The 2022 Directions under section 70B of the Information Technology Act listed defined classes of cyber incidents which must be reported within the relevant period, given the seriousness of some types of incidents, specifically including ransomware events and data breaches.[20]

This closely links incident response and the insurance claims process. For instance, a cyber insurance policy may require prompt notification of the insurer, while the insured may also be subject to a regulatory requirement to notify CERT-In. Therefore, lack of timely reporting may be harmful under both the regulator and the insurance contract. An incident response plan should therefore combine regulatory reporting with notification to the insurer and evidence protection.[21]

Business interruption is another significant coverage issue. Loss is incurred not because the insured technological systems themselves were compromised but because a third-party service provider was blocked. Therefore, the classic line between direct and indirect losses becomes more difficult to maintain in this highly networked digital payments ecosystem.[22]

2.4 Third-Party and Supply-Chain Risk

Most modern fintech firms are not standalone. Their operations rely heavily on cloud platforms, software developers, payment gateways, data centres, cybersecurity providers and other technology vendors. Thus, a cyber incident at one of these vendors can affect a number of downstream entities and customers.[23]

The RBI's 2024 Directions sit squarely on vendor risk and require PSOs to analyse and manage technology and cyber risk concerning third parties. In addition, they recognise the importance of security controls in vendor environments and the need to manage critical service providers.[24]

The above has far-reaching implications for cyber insurance. Usually, cyber policies focus on the insured's own systems. However, the source of disruption may originate in the infrastructure of a vendor. Insurance contracts therefore need to differentiate between direct cyber incidents and those stemming from third parties or supply chain attacks. Contingent business interruption is a key area of concern in this respect.[25]

Table 1: Major Cyber Risks and Potential Cyber Insurance Responses

Cyber Risk	Potential Consequence	Possible Insurance Response	Major Coverage Issue
Phishing and social engineering	Fraudulent transfer	Cybercrime or social engineering cover	Voluntary transfer exclusions
Data breach	Privacy claims and response costs	Data breach and liability cover	Regulatory and third-party liability
Ransomware	System disruption and restoration costs	Cyber extortion and business interruption cover	Exclusions and policy conditions
Payment fraud	Direct financial loss	Payment fraud or crime extension	Authentication and internal controls

[20] Indian Computer Emergency Response Team, Directions under Section 70B of the Information Technology Act, 2000 (Apr. 28, 2022).
[21] *Id.*; Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.
[22] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.
[23] *Id.*
[24] *Id.*
[25] *Id.*; Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

Cyber Risk	Potential Consequence	Possible Insurance Response	Major Coverage Issue
Third-party attack	Service disruption	Contingent business interruption	Causation and vendor dependency
Cloud service failure	Loss of digital services	Business interruption extension	Service provider exclusions
Regulatory investigation	Legal and compliance costs	Regulatory cover where insurable	Scope of coverage

Source: Compiled by the author based on the Indian legal and regulatory framework concerning cybersecurity, digital payments and cyber insurance.

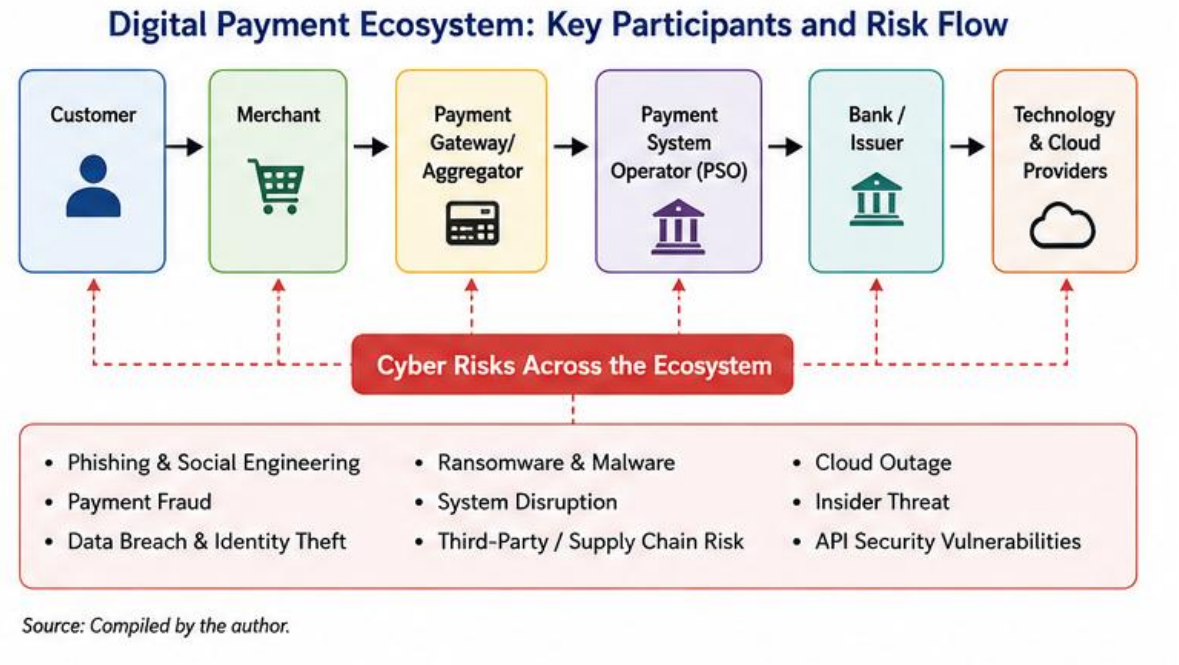

FIGURE 2. *Digital Payment Ecosystem*

3 INDIAN LEGAL AND REGULATORY FRAMEWORK

3.1 Information Technology Act, 2000

The Information Technology Act, 2000 is one of the Indian cyber legal systems covering unauthorised access, damage to computer systems, computer insider crimes and responsibilities of intermediaries among other issues. The Act provides the statutory basis for the legal effects following a cyber risk event.[26]

The current Indian legal framework is relevant to fintech companies beyond criminal liability. Cyber risk issues ranging from unauthorised access and damage to computer resources to data security would challenge fintech companies to evaluate the legal aspects of these risks. The Act is therefore one part of the Indian legal framework for cyber risk.[27]

Cyber insurance does not affect existing statutory responsibilities under the IT Act. The insured party is still liable for adhering to the applicable requirements and the reporting obligations of the IT Act. Cyber

[26] Information Technology Act, 2000, §§ 43, 43A, 66.
[27] *Id.*

insurance simply allows transfer of risks related to the insured events and does not affect the current liabilities of the insured party.[28]

3.2 Payment and Settlement Systems Act, 2007

The Payment and Settlement Systems Act, 2007 serves as the statutory basis for the regulation of payment systems by RBI. The Act duly empowers the RBI to regulate payment systems and issue directions for the same. Hence, it is important to understand the legal obligations of participants in the digital payment ecosystem in India.[29]

The RBI's 2024 cyber resilience directives for non-bank PSOs were published under sections 10(2) and 18 of the PSS Act. They aim to bolster the cyber resilience of payment systems and institute governance and operational processes for cyber risk.[30]

The PSS Act renders certain implications on cyber insurance underwriting. Underwriters of payment operators cannot dismiss exposures to cybersecurity as wholly voluntary risk management practices. The expected adherence in most cases may become a factor in their underwriting review.[31]

3.3 RBI Cybersecurity Framework

The RBI has increasingly imposed cybersecurity requirements on banks and payment ecosystem participants. The 2021 Master Direction on Digital Payment Security Controls set out common minimum-security requirements for covered regulated entities, covering governance, secure application development, threat modelling, security testing, logging and monitoring and secure-by-design principles.[32]

The 2024 Master Directions that are applicable to non-bank PSOs go further, by combining cyber resilience with digital payment security. The Directions cover governance, risk assessment, identity and access management, network security, application security, security testing, vendor risk, data security, incident response, business continuity, cloud security, and payment fraud monitoring.[33]

These can in turn serve as a helpful basis for risk-based underwriting of cyber insurance. Insurers can examine the applicant's authentication controls, monitoring systems, incident handling, backup policies, vendor management, and business continuity protocols. To be certain, however, it should be noted that regulatory requirements do not automatically constitute contractual warranties. The legal effect of a breach would be determined by the terms of the insurance contract and the manner in which those terms are drafted.[34]

A distinction should therefore be made between regulatory compliance and insurance coverage. The application of regulatory standards may have an important role to play in risk quantification but policy wording should delineate those security requirements that are material conditions of coverage and those that offer warranties to the insurer, and the consequences that accrue to the insured on the grounds of breach of the latter.

---

[28] *Id.*; Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.
[29] Payment and Settlement Systems Act, 2007.
[30] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.
[31] *Id.*
[32] Reserve Bank of India, Master Direction on Digital Payment Security Controls, *supra* note 2.
[33] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.
[34] *Id.*; Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

3.4 Data Protection and CERT-In Requirements

Data protection has become an increasingly important part of cyber risk management. The Digital Personal Data Protection Act, 2023, provides a statutory framework for protecting digital personal data. The implementation of the Act is brought about through provisions on commencement and rules made thereunder.[35]

With respect to payment services, data protection regulation is especially important to fintechs. A cyber incident can have multiple privacy, regulatory, contractual and insurance implications. In such scenarios, the insured may need to deal with customers, regulators, law enforcement, business partners and insurers because of the incident.[36]

CERT-In requirements are an important part of the framework. The 2022 Directions issued under section 70B of the IT Act put forward requirements on cybersecurity practices and incident reporting. The official guidance from CERT-In also indicates that where all the requisite information is not available in the reporting period, the available information may be given and more information may be provided later.[37]

Cyber insurance is therefore an important part of the event response strategy of a fintech. Insurance policies include requirements on notification, cooperation and evidence preservation. Therefore, the regulatory notification and the insurer's notification need to be carefully coordinated so that it does not prejudice the insured's regulatory position or insurance claim.[38]

3.5 Role of IRDAI

IRDAI has a key role to play in shaping the insurance framework in which the cyber insurance product is offered. In September 2021, IRDAI issued a Product Structure for Cyber Insurance after thinking about the nature of cyber risks in development and the problem with rigid standardisation in the dynamic environment. The regulator considered that cyber risk is influenced by current legislation, advancements in technology, interconnectivity and new forms of loss.[39]

The flexibility of this approach provides insurers with the ability to develop products according to changing cyber risks. However, too much variation in policy terminology and coverage structures can generate uncertainty for policyholders. Therefore, a fintech enterprise could find that it is difficult to compare policy coverage based solely on premium, as the scope of coverage can vary substantially among insurers.[40]

A preferable solution is not necessarily a complete standardisation of all aspects of cyber insurance policy wording. Instead, India could look to implement standardised definitions for core concepts while maintaining flexibility in the additional coverage. Terms such as cyber incident, data breach, social engineering, payment fraud, business interruption, security failure and third-party service provider could be subject to minimum regulatory terminology.[41]

---

[35] Digital Personal Data Protection Act, 2023; Ministry of Electronics and Information Technology, Digital Personal Data Protection Rules, 2025 (Nov. 14, 2025).

[36] Digital Personal Data Protection Act, 2023.

[37] Indian Computer Emergency Response Team, Directions under Section 70B of the Information Technology Act, 2000, *supra* note 20.

[38] *Id.*; Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

[39] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

[40] *Id.*

[41] *Id.*

4 EMERGING COVERAGE GAPS

4.1 Payment Fraud and Social Engineering

The most significant and emerging coverage gap concerns the losses resulting from social engineering. As digital payment fraud becomes more sophisticated, the deception rather than the technical intrusion is often the crux of the problem. The employee may be instructed through an email that looks like it is from the senior management team or the customer may reveal his login details or authorise the transaction.[42]

The legal position is not always clear cut. If the victim has technically authorised the transaction one could argue that it has not been unauthorised. From a commercial and pragmatic point of view, however, the transaction was based on deception. The classification of the inbound claim as cyber or crime insurance or neither depends on such a distinction.[43]

The regulatory framework must therefore evolve to a more functional rather than technical definition. Coverage should be predicated on the type of insured event rather than on the nature and authenticity of the intrusion.[44]

4.2 Regulatory Compliance and Security Warranties

Cyber policies may contain representations and warranties concerning cybersecurity controls. These may relate to multi-factor authentication, encryption, backups, patch management, privileged access controls, endpoint protection and incident response. Such provisions can assist insurers in assessing and pricing cyber risk.[45]

At the same time security warranties can create uncertainty for insured organisations. A fintech enterprise may satisfy the required controls when the policy is issued but subsequently experience a technical deviation. If the warranty is drafted as an absolute condition of coverage a relatively minor breach could potentially be relied upon to deny a claim even where the breach did not materially contribute to the insured loss.[46]

Insurance law places considerable importance upon the terms of the insurance contract. The Supreme Court has repeatedly allowed that the scope of insurance coverage is determined by the contractual terms and that courts must be careful to scrutinise the language of the policy when determining the insurer's liability.[47]

A balanced regulatory approach should therefore require insurers to identify material cybersecurity representations clearly and explain the consequences of breach. Coverage should not automatically be removed because of a technical deviation that has no material connection with the insured loss unless the policy expressly provides for such a consequence and the term is legally enforceable.[48]

---

[42] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

[43] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

[44] *Id.*

[45] Reserve Bank of India, Master Direction on Digital Payment Security Controls, *supra* note 2; Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

[46] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

[47] Oriental Ins. Co. v. Sony Cheriyan, (1999) 6 SCC 451; United India Ins. Co. v. Harchand Rai Chandan Lal, (2004) 8 SCC 644.

[48] United India Ins. Co. v. Harchand Rai Chandan Lal, (2004) 8 SCC 644; Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

4.3 Third-Party and Systemic Risk

A further major gap arises from third-party dependency. A fintech enterprise may maintain adequate internal cybersecurity controls and nevertheless suffer substantial loss because its cloud service provider, payment gateway or technology vendor experiences a cyber disruption.[49]

The RBI's 2024 Directions recognise this interconnectedness by requiring PSOs to assess and manage risks arising from vendors and other third parties within the digital payment ecosystem. This regulatory recognition should be reflected in cyber insurance contracts through clear provisions dealing with contingent business interruption and dependent business interruption.[50]

Insurance policies should specify whether the relevant third party must be an identified service provider, whether the third-party incident must satisfy the definition of a cyber event and whether geographical restrictions apply. Such clarity can reduce disputes concerning causation and the scope of insured loss.[51]

An even more important risk is that of systemic cyber risk. Cyber incidents that take advantage of a vulnerability present in widely used software or cloud infrastructure can result in simultaneous losses among multiple insureds, thereby creating large accumulation risk for insurers.[52]

The risk of systemic cyber events further supports the need for catastrophe modelling, stress testing and suitable reinsurance arrangements. Very large correlated cyber events could strain insurance capacity and question the assumptions about diversification of risk.[53]

5 INTERNATIONAL COMPARISONS

5.1 European Union

The Digital Operational Resilience Act (DORA) is an important comparative model for India. DORA sets out an elaborate regulatory framework for the provision of digital operational resilience in the financial sector. Covered agencies are required to manage risks arising from information and communications technology (ICT) use, report incidents and maintain resilience measures. DORA also sets out rules that apply to ICT third parties.[54]

Adopting an integrated approach suggests that cyber risk is not strictly an information security problem but that it is also an operational and financial resilience challenge. The relevance of this approach for India is that fintech platforms are becoming increasingly integrated into the financial system.[55]

DORA also emphasizes the importance of ICT third-party risk. Given the relatively limited number of technology providers upon which financial institutions rely, a disruption inflicted on one of these providers may affect more than one institution. The relevance of this principle to cyber insurance policy

[49] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.
[50] *Id.*
[51] *Id.*
[52] *Id.*
[53] *Id.*
[54] Regulation (EU) 2022/2554 of the European Parliament and of the Council of Dec. 14, 2022, on Digital Operational Resilience for the Financial Sector, 2022 O.J. (L 333) 1.
[55] *Id.*

also lies in the fact that the source of an insured's cyber exposure need not arise from within the insured's own organisational perimeter.[56]

India need not copy DORA in a literal sense. However, its holistic view of digital operational resilience has implications for the future design of Indian cyber insurance regulation. Cybersecurity compliance, vendor management, incident reporting, and risk transfer via insurance act increasingly in a synergistic manner to contribute to financial resilience.[57]

5.2 United States

The United States offers a divergent paradigm in which cybersecurity regulation is enforced through federal and state level requirements and sector-specific regulatory bodies. New York's cybersecurity regulation for financial services entities offers an important illustration of detailed organisational IT security satisfying requirements that apply to regulated entities.[58]

The US cyber insurance market has also evolved sophisticated underwriting practices. Insurers make increasing use of detailed questionnaires, technical reviews and security information to assess the risk profile of applicants. Policies may include first-party and third-party coverages with extensions covering social engineering, business interruption, ransomware and regulation-related exposures.[59]

The US experience illustrates both the merits and limits of a market-based cyber insurance regulatory framework. On the one hand, flexible underwriting may motivate ultimate policyholders to improve their security posture as robust security measures may affect the availability and pricing of coverage. On the other hand, considerable variation in policy wording can lead to uncertainty as to the precise scope of protection.[60]

India could adopt a hybrid market-regulatory approach. Defining a baseline set of regulatory requirements for coverages and disclosures could permit insurers to maintain some flexibility in the design of policies tailored to different categories of risk associated with fintech and digital payments products and services.

---

[56] *Id.*

[57] *Id.*; Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

[58] N.Y. Comp. Codes R. & Regs. tit. 23, § 500.

[59] *Id.*

[60] *Id.*

Comparative Regulatory Approaches to Cyber Resilience

Aspect	India	European Union (DORA)	United States
Regulatory Nature	Sectoral (RBI, IRDAI, CERT-In, DPDP Act)	Comprehensive EU-wide framework	Sectoral + State-based (Regulatory patchwork)
Focus Areas	Cybersecurity, Payment Security, Data Protection	ICT Risk, Incident Reporting, Resilience Testing, Third-party Risk	Cybersecurity, Standards, Disclosure, State Regulations
Third-Party Risk	Addressed by RBI (2024 Directions)	Strong emphasis with direct oversight of critical ICT providers	Addressed through contracts and supervisory guidance
Incident Reporting	CERT-In (within 6 hours) + Sector Regulators	Harmonised thresholds and timelines	Varies by sector and jurisdiction
Insurance Linkage	Evolving integration with regulation	Operational resilience expected to influence insurance	Driven by market, underwriting practices
Approach	Building coordinated framework	Integrated, principles-based with prescriptive elements	Market-led with regulatory oversight

Source: Compiled by the author.

FIGURE 3. *Comparative analysis*

6 RECOMMENDATIONS FOR LEGAL AND REGULATORY REFORM

Greater consistency between insurance products and clear definitions for common terminology will be instrumental in the development of this space. IRDAI should launch a regulatory glossary which will codify minimum terms and definitions for common terms such as cyber incident, data breach, ransomware, social engineering, payment fraud, business interruption, system failure and third-party cyber event. This will enhance comparability between insurers providing cyber insurance while also allowing them freedom on policy wording.[61]

Underwriting decisions should become more risk-based and include a review of the Internet service provider (ISP) and hosting details supplied by the applicant. Insurers should evaluate the apposite cybersecurity measures implemented by the applicant, which should reflect the regulatory classification and risk profile of the business. The measures are expected to include but not be limited to, adherence to RBI requirements, authentication, security testing, incident response, backup, vendor management, and business continuity.[62]

Inclusion of social engineering and payment fraud coverage in policy wording has received great attention. Insurers should be transparent on whether the fraudulent transfer caused by deception is covered and under what circumstances. This will reduce ambiguity on whether a payment fraud is covered as an unauthorized transaction, a cyber event or a crime event.[63]

Security warranties should also be proportionate and causally connected to the insured loss. Insurers should be able to protect themselves against material cybersecurity deficiencies but a technical deviation

[61] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.
[62] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.
[63] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

that has no meaningful relationship with the loss should not automatically result in denial of an otherwise valid claim unless the policy clearly provides for that result.[64]

Third-party and supply-chain risks should receive explicit treatment. Cyber insurance policies for fintech organisations should address cloud providers, payment processors, software vendors and other critical service providers. The policy should identify when an incident originating with a vendor constitutes an insured event and whether contingent business interruption coverage applies.[65]

Greater coordination between regulators is also necessary. RBI, IRDAI, CERT-In and data protection authorities have distinct institutional responsibilities but their requirements may overlap during a cyber incident. Coordination concerning incident reporting, evidence preservation, investigation, business continuity and insurance claims would reduce regulatory uncertainty and improve incident response.[66]

Insurers should also develop stronger capabilities for assessing systemic cyber risk. As financial technology infrastructure becomes increasingly concentrated a single vulnerability may affect numerous organisations simultaneously. Scenario analysis, catastrophe modelling and appropriate reinsurance arrangements should therefore form an increasingly important part of cyber risk underwriting.[67]

Finally, fintech organisations need to consider cyber insurance as a part of a more comprehensive cyber risk management approach. Insurance should not serve as a substitute for investments in cybersecurity measures. Organisations should establish incident response strategies that bring together technical teams, senior management, legal advisers, the regulators and insurers. This should help mitigate cyber incident impact as well as disputes over insurance coverage.[68]

7 FINDINGS AND DISCUSSION

The analysis shows that the fintech and digital payments ecosystem in India is evolving while the cyber risk environment is also becoming increasingly complex. While the RBI has set out detailed cybersecurity requirements for payment system participants and the CERT-In has set out reporting obligations, the IRDAI has concurrently formulated standards for cyber insurance products. The existence of these regulatory initiatives indicates that cyber risk is now an issue that extends beyond technology and permeates financial and operational resilience.[69]

The central finding of the research is that India's major challenge is not a lack of comprehensive regulations. Rather, the challenge is dealing with the fragmentation of regulations on cybersecurity, regulations on payment systems, data protection laws, and insurance laws. As a result, a fintech firm may have

[64] United India Ins. Co. v. Harchand Rai Chandan Lal, (2004) 8 SCC 644.

[65] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

[66] Indian Computer Emergency Response Team, Directions under Section 70B of the Information Technology Act, 2000, *supra* note 20; Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

[67] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

[68] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

[69] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1; Indian Computer Emergency Response Team, Directions under Section 70B of the Information Technology Act, 2000, *supra* note 20.

multiple obligations relating to a single cyber incident and the insurance policy may use distinct terminology and differing contractual requirements.[70]

A second finding is the increasingly difficult line between cyber risk and financial crime. Social engineering is a clear example. A spoofed digital payment can be a technology-based incident, a human-enabled fraud and a financial crime simultaneously. Insurance products based on a rigid separation between these may create coverage uncertainty.[71]

A third finding is third-party dependency. Fintechs often use third-party technology vendors for cloud services, software development, payment systems and other critical services. Cyber insurance in the Indian context should not only evaluate the cyber risk of the insured entity but also material third-party-dependent risk.[72]

A fourth finding is that cybersecurity compliance with regulatory requirements can become an important underwriting criterion. RBI governance, security testing, vendor risk governance, data security, incident response and business continuity requirements offer a good starting point to assess cyber risk. However, cybersecurity compliance requirements should not be automatically treated as pure and absolute insurance warranties without careful examination of causation and fairness in contractual terms.[73]

The comparative analysis suggests that India can learn from international experience without directly copying foreign models. The EU approach demonstrates the value of integrated digital operational resilience while the US experience highlights the strengths and weaknesses of flexible market-based cyber insurance. India can incorporate these lessons while retaining a regulatory framework suited to the particular characteristics of its domestic digital payment ecosystem.[74]

8 Conclusion

As India's fintech and digital payment ecosystem continues to grow, cyber risk insurance will become increasingly important. The scale, technological interconnectedness and speed of digital financial services mean that one cyber incident can generate losses extending well beyond the immediate technological impact. Payment fraud, social engineering, data breaches, ransomware, business interruption and third-party technology failures may generate financial, regulatory, contractual and reputational consequences simultaneously.[75]

India has built essential components of a cyber risk governance structure with the Information Technology Act, 2000, Payment and Settlement Systems Act, 2007, Digital Personal Data Protection Act, 2023, RBI cyber security directions, CERT-In mandates and the IRDAI framework for cyber insurance.

---

[70] Information Technology Act, 2000; Payment and Settlement Systems Act, 2007; Digital Personal Data Protection Act, 2023; Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

[71] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

[72] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

[73] Reserve Bank of India, Master Direction on Digital Payment Security Controls, *supra* note 2; Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

[74] Regulation (EU) 2022/2554; N.Y. Comp. Codes R. & Regs. tit. 23, § 500.

[75] Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

RBI's 2024 Master Directions are particularly noteworthy because they bind together cyber resilience and cyber risks to digital payment security and third parties.[76]

Yet many gaps remain. Payment fraud and social engineering underline the challenge of separating cyber incidents from financial crimes. Security warranties can breed doubt where contract stipulations are strictly more demanding than the causal link between a cybersecurity failure and the insured loss. Third-party connectivity bears further doubt over contingent business interruption while systemic cyber events give rise to difficult issues about insurance capacity, accumulation and reinsurance.[77]

Consequently, the future development of cyber risk insurance in India should concentrate on regulatory coordination. It is not about augmenting the number of regulations. Instead of merely clarifying descriptive terminology and policy disclosures, IRDAI can facilitate transparency and information disclosure facilitating decision making and lessening asymmetric information between insurers and insured. The concomitant advancements in standard-setter mechanisms such as cybersecurity best practice standards promulgated by RBI, and incident reporting obligations enforced by CERT-In, can provide enhanced foundations for the assessment of cyber risk.[78]

A mature cyber insurance market in India should pursue three interlinked objectives: efficient transfer of cyber risk, enhanced incentives for better cybersecurity, and increased legal predictability. The requisite regulatory architecture must be cognisant of the fact that financial technology and cyber risk insurance share such strong interdependencies with cybersecurity best practice, payment system safeguards, and data protection norms.[79] To that end, an integrated regulatory framework that fosters consumer confidence, ensures financial system stability, and promotes sustainable growth of digital financial services can bring about subtler but more important gains for the Indian economy.

* * *

[76] Information Technology Act, 2000; Payment and Settlement Systems Act, 2007; Digital Personal Data Protection Act, 2023; Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1; Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3.

[77] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3; Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

[78] *Id.*; Indian Computer Emergency Response Team, Directions under Section 70B of the Information Technology Act, 2000, *supra* note 20.

[79] Insurance Regulatory and Development Authority of India, Product Structure for Cyber Insurance, *supra* note 3; Reserve Bank of India, Master Directions on Cyber Resilience and Digital Payment Security Controls for Non-bank Payment System Operators, *supra* note 1.

Law in the Digital Decade: Evidence, Intellectual Property and Markets, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026) · Open access under CC BY-NC 4.0 · https://doi.org/10.63108/VAB.LDD.2.19

CHAPTER 19

The Janus Paradox: Evaluating Tax Friction, Regulatory Blind Spots and Capital Flight Under the Indian Virtual Digital Assets Regime

TARUN KUMAR SRIVASTAVA[*] AND KESHAV[**]

Virtual Digital Assets (VDA) are constructed using cryptography, exist in intangible form, and are identified by incorporating an electronic document into the electronic transaction record, with commercial significance. The inclusion of cryptographic technology into the global financial regime has formulated a 'decentralised' system that challenges the traditional legal framework. VDAs' autonomous operating capability enables them to serve as a system protected by an encryption algorithm, for both a virtual accounting system and a medium of exchange. This study highlights and examines the structural regulatory gap in the statutory framework governing VDAs, as reflected in the Finance Act and the Prevention of Money Laundering Act. While the statutory provisions and Financial Intelligence Unit-India (FIU-IND) reporting directives treat VDAs primarily as a subject of revenue collection and transaction monitoring, this paper, by applying a mixed-method methodology combining doctrinal statutory implications of tax and Anti-Money Laundering (AML) framework with scientific market volume implications in the Indian regime, analyses the operational and practical impacts of these provisions. This paper demonstrates that stringent fiscal measures, especially a 30% flat tax without a loss set-off and a Tax Deducted at Source (TDS) of 1% on the gross transactional value, impact liquidity in the domestic market, creating a 'tax leakage' by trading a significant amount into offshore platforms, undermining the enforcement goals of FIU-IND. This paper also explores the jurisdictional gap of self-hosted wallets and automated smart contracts, which is created through the reliance of PMLA on centralised reporting entities. This paper suggests a statutory framework to establish property rights by changing the 1% TDS regime to 0.01% to conserve transaction tracking without impacting market liquidity, enable same-class loss set-offs, and enforce forensic monitoring to address the cryptographic architecture.

Keywords: Virtual Digital Assets, Cryptographic Technology, Finance Act, Prevention of Money Laundering Act, Tax Leakage.

1 INTRODUCTION

Virtual Digital Assets (VDAs), also called 'electronic assets', are constructed using cryptography, exist in intangible form, and are identified by incorporating an electronic document into the electronic transaction record, with commercial significance.[1] VDAs incorporate virtual goods and services, tokens and cryptocurrencies, which create difficulties and problems for the Indian economy and monetary canon.[2] In India, the use and influence of VDAs increased significantly after the Supreme Court (SC) set aside the circular of the Reserve Bank of India (RBI) of 2018, by which it directed regulated financial institutions to cut off access to VDAs, and held that, as there is no statutory framework enacted by the

[*] 4th Year B.A. LL.B. (Hons.) Student at Central University of South Bihar, Gaya, Bihar, India.
[**] 4th Year B.A. LL.B. (Hons.) Student at Central University of South Bihar, Gaya, Bihar, India.

[1] Income Tax Act, 1961, § 2(47A) (inserted by Finance Act, 2022).

[2] *Id.*; Ministry of Finance, Reply to Lok Sabha Starred Question No. 10 (Cryptocurrency) (July 18, 2022).

Parliament that prohibits the use and access of VDAs, the executive body cannot deny the use of monetary and commercial assets of any form to the citizen as it violates the citizens' right of freedom, which a person has, to carry on his business which is incorporated under Article 19(1)(g)[3] of the Indian Constitution.[4] The Parliament, in 2022, enacted the Finance Act[5] by amending the Income Tax Act[6] to impose tax on the income generated through VDAs, where the Parliament introduced Section 115BBH[7] to impose a flat tax of 30% on profits earned without allowing intra-asset loss offset, and also included Section 194S[8], which imposes 1% tax on gross transactions. The central government also increased the scope of Section 2(1)(sa)(vi)[9] of the Prevention of Money Laundering Act by including Virtual Asset Service Providers (VASPs), including custodial service providers and transfer-intermediary applications. Accordingly, digital asset middlemen are specified as 'reporting entities' dependent on customer due diligence (KYC), business registration compliance, and compulsory filing of Suspicious Transaction Reports (STRs) with the Financial Intelligence Unit-India (FIU-IND).[10] All these provisions act as a tool for the supervision and monitoring of virtual digital assets, but all these provisions have a deficiency in any civil consumers' remedies and primary ownership rights, as the current statutory provisions lack codification in the Indian legal system.

2 BACKGROUND

2.1 Conceptual Background

In the Indian legal system, the Virtual Digital Asset (VDA) is outlined and interpreted in section 2(47A)[11] of the Income Tax Act, as a cryptographically created cypher, numerical code, or emblem enabling a virtual depiction of value that can be transmitted, stored, or traded electronically, expressly excluding fiat currencies. In the contemporary system, these assets operate in two parallel models: a centralised, integrated system and a decentralised, distributed system. The centralised and amalgamated system mimics the non-contemporary financial system, which includes sustained control over its consumers' private cryptographic keys and enables customers to manage domestic off-chain ledgers to assist fiat-to-crypto liquidity.[12]

By contrast, decentralised, distributed, self-hosted systems like wallets enable transfer and transaction on a peer-to-peer basis in a distributed system, which enables customers to access and possess their own private keys and eliminate tech intermediaries. Any transmissions among the customers are governed through 'smart contracts', which include predetermined, self-enforcing software code which is based on a 'public blockchain' that automatically resolves the transfer of tokens if the system's mathematical calculations are met, totally sidestepping human intervention. Whereas the centralised mechanism operates within a traditional statutory framework enabling the role of intermediaries, the decentralised system

[3] INDIA CONST. art. 19, cl. 1(g).
[4] Internet & Mobile Ass'n of India v. Reserve Bank of India, AIR 2020 SC 298.
[5] Finance Act, 2022.
[6] Income Tax Act, 1961.
[7] *Id.* § 115BBH.
[8] *Id.* § 194S.
[9] Prevention of Money Laundering Act, 2002, § 2(1)(sa)(vi).
[10] Advait Sharma, *Regulation of Virtual Digital Assets Under PMLA in India*, 7 INT'L J. MGMT. & HUMAN. 112, 115–20 (2024).
[11] Income Tax Act, 1961, § 2(47A).
[12] Kumar Rishav & Divya Venugopalan, *The Digital Era: Cryptocurrency and Its (In)efficient Taxation and Regulation Policy*, 2 HPNLU J. TAX'N L. 91, 95–99 (2023).

executes on the doctrine of smart contracts, opposing the conventional framework by operating without any corporate intermediaries.[13]

2.2 Historical Background

In the Indian legal system, VDAs, their concept and their regulatory approach have evolved continuously over a period of time. The VDAs were introduced in the early 2010s in India, which was a new concept for the general public as well as for the government.[14] Between 2013 and 2018, the Reserve Bank of India (RBI) adopted a cautionary and admonitory stance toward VDAs, where the RBI issued public advisories warning the general public against the monetary, functional, and statutory risks of VDAs.[15] Further, in 2018, the RBI implemented a constructive 'ring-fencing' framework, which forbade regulated banking enterprises from rendering any VDAs and cryptographic asset-related services to users and customers. In 2020, the ban on VDAs and other cryptographic assets was challenged in court, and the court struck down the ban and held that the complete administrative ban on VDAs and other cryptographic assets without any scientific or latent harm violates Article 19(1)(g)[16] of the Indian Constitution.[17] In the aftermath of the judicial precedent, the Parliament enacted the Finance Act and formulated the tax regime, in which it introduced an imposition of a 30% tax on profits and an imposition of a 1% tax on gross transactions.[18] Ultimately, in 2023, the central government incorporated the concept of VDAs and other cryptographic assets into the provisions of the Prevention of Money Laundering Act (PMLA) statutory framework, where it made compulsory the filing and reporting in conformity with FIU-IND.[19]

2.3 Theoretical Background

This paper evaluates the administration of VDAs in the Indian legal and administrative system by evaluating three core principles. First and foremost is the Janus Model of administration, which asserts the push and pull of the act of the state and the objective of the regulation. The Parliament has imposed a flat tax of 30% on profits earned by the customers of VDAs and a tax of 1% on gross transactions, where the Parliament has adopted restrictive measures aimed at direct control of VDAs, and restricts the trading of VDAs in the Indian fiscal system. This measure of Parliament directly opposes the objective of the anti-money laundering motto of the state, where the main objective is the decentralised regulation of the market, and supports the principle of *laissez-faire*. Both the objective and statutory framework of the state are a paradox (Janus) in themselves, where both frameworks are in conflict with each other but also co-exist with each other.[20]

Second is the principle of 'Economic Analysis of Law', which asserts that when the tax slab and its regulatory compliance are greater than the acceptable length, it leads to tax evasion and the transfer of money from Indian territory into offshore accounts and non-compliant locations.[21] Third is the doctrine of 'code is law', which was developed by Lawrence Lessig, which asserts the legal non-alignment between

---

[13] Kevin Werbach, *Trust but Verify: Why the Blockchain Needs the Law*, 33 BERKELEY TECH. L.J. 478, 501–12 (2018).
[14] Rishav, *supra* note 12, at 100.
[15] Shreya Singh, *Cryptocurrencies: Unravelling Regulatory Realities and Tax Framework*, 4 INDIAN J. INTEGRATED RSCH. L. 1, 3–6 (2024).
[16] INDIA CONST. art. 19, cl. 1(g).
[17] Internet & Mobile Ass'n of India v. Reserve Bank of India, AIR 2020 SC 298, ¶¶ 220–225.
[18] Vikash Gautam, *Virtual Digital Asset Tax Architecture in India: A Critical Examination* 6 (Esya Centre Special Issue No. 208, 2023).
[19] Sharma, *supra* note 10, at 115, 120–22.
[20] Rishav, *supra* note 12, at 101–03.
[21] Singh, *supra* note 15, at 10–11.

the statutory framework and the contemporary decentralised, automatic, systematic algorithm.[22] The conventional statutory framework asserts the presence of an intermediary, such as a corporate entity or a natural person, who can be held liable for any breach; the contemporary decentralised algorithm operates on the concept of smart contracts and localised wallets, where there is little involvement of any intermediaries. This establishes a statutory gap where the legal provisions have not evolved in line with the evolution of the algorithm of the tax regime.

3 OBJECTIVE OF THE STUDY

This paper explores the statutory rules under section 2(47A)[23], section 115BBH[24] and section 194S[25] of the Income Tax Act. This paper articulates the PMLA rules of reporting and how they are applied to the contemporary VDAs and other cryptographic algorithms. This paper examines the transmission of currency from Indian trading and monetary platforms to foreign offshore accounts and destinations. This paper analyses and identifies the gap in the statutory framework, which includes the property and possession rights of an individual, the safety of investors and the decentralised algorithm of e-wallets. This paper suggests a balanced approach for new statutory and planning reforms, which is appropriate for the contemporary Indian tax regime.

4 RESEARCH METHODOLOGY

This paper utilises a mixed Theoretical-Practical approach, where this paper analyses the enacted legal provisions of the Income Tax Act, PMLA Rules, parliamentary notifications, administrative notifications, and precedents, with research data on market quantity response to the novel statutory provisions of tax. The primary source of data includes the Income Tax Act, Finance Act, PMLA and precedents of the court, and the secondary source includes reports of think tanks, academic articles, etc. This paper relies on market evidence, as the tax documents of the individual and the details of foreign accounts and destinations cannot be accessed.

5 ANALYSIS AND FINDINGS

5.1 The Tax Regime and Its Application

The Indian tax statutory regime has incorporated provisions in the Income Tax Act that have an impact on the Indian fiscal market and its structure. Section 2(47A) has endorsed an extensive, tech-driven explanation incorporating any virtual goods and services, tokens derived through the process of cryptography, which fails to differentiate between a monetary token and its operative usefulness or administrative possessions.[26] Section 115BBH imposes a flat tax of 30% on profits earned, with strict and rigid statutory provisions that disallow the deduction of any expenditure other than the cost of

[22] LAWRENCE LESSIG, CODE: VERSION 2.0, at 1–15 (2006).
[23] Income Tax Act, 1961, § 2(47A).
[24] *Id.* § 115BBH.
[25] *Id.* § 194S.
[26] Mayank Jain & Tanya Saraswat, *Deciphering Section 2(47A): The Overbreadth of Virtual Digital Asset Definitions Under Indian Tax Law*, 9 NAT'L L.U. DELHI L. REV. 118, 122–26 (2023).

acquisition, including intermediary fees or blockchain operational costs, and also does not allow intra-asset loss offset, where a gain in one transaction offsets a loss in another transaction. This statutory framework imposes a great burden on the taxpayers and other market players. Take a market investor who gains a profit of Rs. 30,000 in Token 1 and subsequently incurs a loss of Rs. 30,000 in Token 2. Yet, the market investor is liable to pay 30%, i.e. Rs. 9,000 (Rs. 9,360 with the 4% health and education cess), as a direct tax. Moreover, Section 194S states a compulsory responsibility to deduct 1% as Tax Deducted at Source (TDS) on the gross revenue of each transaction. This framework imposes a burden upon the market users, wherein a market investor secures a margin of 0.20% in his transaction, and the statute imposes a burden of 1% TDS. This leads to the depletion of venture assets, making the market considerably less attractive for investment.

5.2 Market Movement

The enactment of the Finance Act had a significant impact on the market and its reorganisation. Within three months of its implementation, the everyday collective business volumes decreased by 90%,[27] where it came from transactions of millions of dollars to a negligible level. This phenomenon led to the event of 'offshoring', where, in reality, the demand of the market did not reduce, but the allocation and investment of capital in the form of profit was shifted from the Indian fiscal market to offshore markets and destinations. A business volume of $3.8 billion was transmitted to the offshore destinations, which has not only impacted the capital market but also the tax collection of the Indian fiscal regime.[28] Moreover, the 'on-chain' measures specified that there is a prominent shift from highly regulated intermediaries and obligations to autonomous e-wallets, which are privately held and do not impose high obligations upon their users. The shift to decentralised algorithms has led to the flow of money outside the range of the domestic banking system, which has affected the whole function of lending and investment by banks, individuals or organisations into small businesses, enterprises, or any developmental project or infrastructure.

5.3 The Implementation and Restriction of PMLA

The Prevention of Money Laundering Act was enacted by Parliament to prevent money laundering and encourage the free flow of the economy and investment in the market. In March 2023, the central government notified VDA service providers as 'reporting entities' under section 2(1)(sa)(vi)[29] of the Prevention of Money Laundering Act, where there were mandatory provisions for Know Your Customer (KYC), preservation of a five-year transaction register and filing of Suspicious Transaction Reports (STRs) with the Financial Intelligence Unit-India (FIU-IND). The statutory framework of PMLA has three main issues. First and foremost is the 'intermediary issue', where the conventional structural framework of PMLA includes a structural middleman that can be held accountable for any breach of conduct. But, in the contemporary transactional framework like the peer-to-peer (P2P) model, where a transaction is concluded between two people without the presence of any intermediary, like when a person transfers

[27] *Id.* at 120.
[28] Gautam, *supra* note 18, at 4, 22 (cumulative trade volume of around USD 3,852 million shifted from Indian to foreign centralised VDA exchanges during Feb.–Oct. 2022).
[29] Prevention of Money Laundering Act, 2002, § 2(1)(sa)(vi).

a bitcoin from their hardware to another's, there is no presence of any unified structure to maintain KYC or to follow the FATF travel rule. This escapes the ambit of the PMLA statutory framework.

Second is that Section 17[30] of the PMLA provides for the search of premises and the seizure of records, on which marks of identification may be placed, but the section cannot be implemented effectively as the VDAs operate on a decentralised monetary policy with smart automatic contracts, where decentralised exchanges such as Curve or Uniswap operate on self-performing algorithms and lack any centralised machinery with head offices, staff, board members, directors, or any personnel who could be held liable for any non-compliance or breach of the procedure.

Third is the transnational friction of jurisdiction, where some Indian users route transactions through offshore platforms that lie outside the reach of the Indian tax regime, as the foreign destination does not maintain any local jurisdictional existence, which leads to great difficulty in tracking the path of flow of the transaction and maintaining a real-time register. While the central government has taken measures like blocking of URLs and offshore mobile accounts and applications by the Ministry of Electronics and Information Technology (MeitY), it can be circumvented through VPNs and similar tools.

6 POLICY SUGGESTIONS

The contemporary Indian tax regime is based on the conventional method of statutory framework, which leads to the phenomenon of Janus Paradox, where the Indian tax regime imposes high tax slabs with a strict compliance mechanism with the objective of promoting the idea of a free and open market for investment. To resolve the issue of the Janus Paradox in the Indian tax regime, there are four key suggestions that should be implemented in the existing statutory framework. First and foremost is the enactment of separate legislation for VDAs, which must define the features of theoretical tokens separately from operational service tokens and also establish a protective statutory framework for the maintenance of users' assets in the event of a breach of security, etc.

Secondly, Parliament should amend the statutory provision of Section 194S[31] of the Income Tax Act and change the slab of TDS from 1% to 0.01%. A market businessman trades Rs. 1 crore in the market and therefore, according to the current framework, he is liable to pay Rs. 1 lakh, which could somehow deplete the capital from the market. A slab of TDS of 0.01% can give rise to the liability for the market businessman to pay Rs. 1,000, which ultimately protects liquidity in the market, as the businessman could utilize the remaining amount, i.e. Rs. 99,99,000, for further investment in the Indian market, and will also increase the number of investors investing in the Indian market because of lesser TDS slab compared to other countries, which will ultimately increase the overall TDS collection.

Thirdly, intra-class loss offsets must be incorporated in Section 115BBH(2)(b)[32] of the Income Tax Act, where if a person profits Rs. 50,000 in token 1 and incurs a loss of Rs. 50,000 in token 2 of the same valuation, then in this case he should be allowed to offset his tax, which is not permitted in the current tax regime, where it violates the doctrine of ability-to-pay by generating a direct cess on monetary-neutral income.

[30] *Id.* § 17.
[31] Income Tax Act, 1961, § 194S.
[32] *Id.* § 115BBH(2)(b).

Lastly, section 115BBH(2)[33] of the Income Tax Act should be amended to allow a deduction for the fees of the blockchain algorithm network, where currently it disallows the deduction of mandatory fees for accessing the blockchain system. Consequently, the government should include the supervision of on-chain (blockchain) and other cryptographic networks under FIU-IND, where currently it can only supervise tangible intermediaries.

7 CONCLUSION

The contemporary Indian tax regime is based on the conventional method of statutory framework, which leads to the phenomenon of Janus Paradox, where the Indian tax regime imposes high tax slabs with a strict compliance mechanism to encourage the object of a laissez-faire market for investment. To deal with VDAs as a form of income, the state has taken several measures, such as the imposition of a flat 30% tax on profit, 1% TDS on each transaction, inclusion of VDA service providers in FIU-IND reporting under PMLA, which are somewhat ineffective in properly achieving the objective of a free and open market for investment, which has resulted in the reduction in business quantity in the Indian market by 90%. To deal with the above issue, this paper proposes a reduction in the TDS slab from 1% to 0.01%, allowing for offsets in the event of losses incurred by investors, and lastly, the adoption of on-chain monitoring in the current FIU-IND regime. All the above measures will increase the flow of liquidity as well as the number of investors in the Indian market.

* * *

[33] *Id.* § 115BBH(2).

About the Publisher

VidhiAagaz is an academic publisher based in Maihar, Madhya Pradesh, publishing since 2017. It is a member of Crossref (DOI prefix 10.63108), and its publications are indexed in prominent databases, including HeinOnline, Google Scholar and Manupatra. Every chapter of this book carries its own DOI and can be read free of charge at vidhiaagaz.com.

Publish your conference with us

Universities, departments and conference organisers can publish their conference papers with us as an edited book or as a special issue of our journals, the *International Journal of Law Management & Humanities* (IJLMH) and the *International Journal of Legal Science and Innovation* (IJLSI). We take a volume from manuscript to print: copy-editing, citation checks, typesetting, ISBN and DOI registration, open-access publication online, and printed copies.

Write to support@vidhiaagaz.com

vidhiaagaz.com/collaborate

Law in the Digital Decade

Evidence, Intellectual Property and Markets

Digital technology has changed what counts as evidence, what can be owned and how markets behave. Courts now weigh certificates for electronic records, deepfakes and traces left in the metaverse. Copyright law is asked who owns what a machine generates and whether training a model on protected works is lawful. Regulators confront dark patterns, algorithmic management of gig workers, artificial intelligence in the boardroom and the taxation of virtual digital assets.

This volume brings together nineteen chapters from the CLRS International Conference on Law in the Digital Decade, held in September 2026, in three parts: digital evidence, cybercrime and the criminal process; intellectual property and generative AI; and markets, corporations and the digital economy. Reading the Bharatiya Sakshya Adhiniyam, 2023, the Copyright Act, 1957 and India's consumer, corporate and tax law beside comparative experience, the chapters identify gaps and propose reforms.

The book is written for students, researchers, practitioners and policymakers.

About the Editors

Gyan Prakash Kesharwani is the Chairperson and Publishing Editor at VidhiAagaz, India.

Prasanna Kumar Shukla is a Doctoral Researcher at National Law Institute University, Bhopal, India.

VidhiAagaz®
Inking Your Brain

www.vidhiaagaz.com

ISBN 978-81-181981-0-9